Aberystwyth University

3D Surface Texture Analysis of High Resolution Normal Fields for Facial Skin Condition Assessment

Supervisors:

Author:

Bernard Tiddeman

bpt@aber.ac.uk

Alassane Seck

als31@aber.ac.uk

Hannah Dee

hmd1@aber.ac.uk

A thesis to be submitted for the degree of

Doctor of Philosophy

October 2016

Abstract

In this thesis, we introduce new methods for analysing three dimensional surface texture using high resolution normal fields and apply these to the detection and assessment of skin conditions in human faces, specifically wrinkles, pores and acne. The thesis is part of a project sponsored by Unilever with an interest in applying the outcome of the research to facial skin care product development. This explains our focus on facial skin conditions.

The main contributions of this thesis are the introduction of three methods of extracting texture descriptors from high resolution surface orientation fields, a comparative study of two-dimensional and three-dimensional skin texture analysis and the collection of an extensive dataset of high resolution 3D facial scans presenting various skin conditions. The dataset includes human rating judgements on the presence of certain skin conditions.

Computer aided skin condition assessment has been mostly addressed using two dimensional texture analysis techniques on skin images or coarse geometrical features extracted from the skin's three dimensional macro structures. While the first trend ignores the three dimensional nature characterising most of the skin conditions, the latter mainly deals with geometrical features that are not fine enough to capture skin structures at the meso and micro scales.

Advances in three dimensional surface imaging during the last few decades brings the possibility of capturing human skin's fine geometrical structures and reflectance properties with unprecedented quality and resolution (down to the level

of the pores). The methods proposed in this work aim at leveraging these advances and revisit the formulation of texture analysis as a three dimensional problem.

For data collection we set up a Lightstage to capture high resolution facial normal fields along with reflectance properties. The collected data are photo-realistically rendered and presented to the general public for annotations indicating the presence of the studied skin conditions. These constitute the ground truth on which we apply the proposed methods and learn models for detecting and assessing facial skin conditions.

We also demonstrate that some of these three dimensional surface texture descriptors can be extended to synthesize highly detailed skin structures and simulate the studied skin condition on normal faces.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors Dr Bernard Tiddeman and Dr Hannah Dee for their continuous support throughout my PhD research. Without their knowledge of the field, guidance and patience, this thesis would have not seen the light of day.

Also, I would like to thank Unilever and the Department of Computer Science of Aberystwyth University for funding this PhD and for having provided me with resources and very suitable environment to conduct my research.

Thanks to my fellow office mates, in particular Alexandros Giagkos for his help, advices and friendship.

Finally, I would like to thank my family: parents, brothers, sisters and friends for all their support in my life in general and throughout my PhD in particular.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Objective

The main objective of this work is to develop 3D surface texture analysis methods for classifying facial skin condition such as wrinkles, pores and acne (Figure 1.1). The project is sponsored by Unilever with an interest in skin care applications, which justifies our focus on facial skin conditions.

Previous work on computer based skin analysis has mainly focused on 2D texture descriptors with a considerable emphasis on facial wrinkle detection and age classification [46, 58]. 3D geometrical features have been used in studies on detecting skin cancer [44, 86]. However these features are generally extracted from macro structures and do not exploit the skin fine texture.

The 3D surface texture descriptors we propose operate directly on the geometrical structure of the skin at different scales through multi-resolution analysis schemes,

thus capturing the skin disruption in the macro scale (acne, wrinkles) as well as in meso and micro scales (fine wrinkles, pores).

The analysis is performed on facial skin normal fields. These are acquired using photometric stereo techniques which are proven to be effective for capturing highly detailed skin geometrical properties – down to the level of the pores – in the form of surface normal orientations [52, 53].

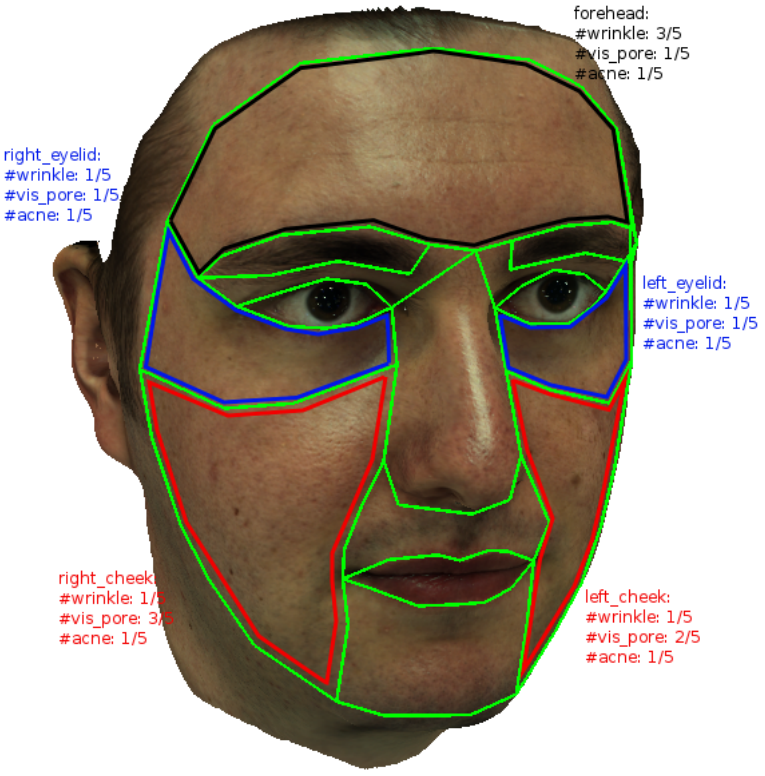


FIGURE 1.1: Ojectives I : 3D skin texture analysis

1.2 Motivations

The main motivation of this work is to explore and evaluate the value added by introducing 3D geometrical properties on rough surface texture analysis, and

more particularly on human facial skin.

In general, surface texture can be defined as the structural variation in appearance of the surface due to its reflectance or 3D geometrical properties. For a long-time the 3D geometrical aspects of texture have been neglected in computer vision, posing texture characterisation as a 2D problem. It is obvious that surfaces that involve variational relief, like wrinkly skin, will present different apparent textures on a 2D image according to the imaging conditions (light direction and viewpoint), due to 3D properties.

Around 20 years ago, a new trend aiming to address this three dimensional nature of texture emerged. With the work of Dana et al. [12] and Leung & Malik [48], these approaches extract texture descriptors from a number of images of the same surface under different illumination conditions and viewing directions. These methods have two implications:

A considerable number of imaging conditions are required to capture all the fine 3D geometrical properties of the measured surface (for example in the Columbia-Utrecht dataset, 200 different combinations of viewing and illumination directions are used [12]).

The second implication is a direct consequence of the first and concerns the texture characterisation itself: the resulting feature spaces present an extremely high dimensionality which leads to unavoidable compression, with underlying data loss issues. For example, to model a BTF (Bidirectional Texture Function),

which is one of the most accurate ways used to represent illumination/view independent texture, one needs to compute a reduced and parametric representation from a dataset that can be hundreds of gigabytes in size. The accuracy of the resulting model generally depends on the quantisation method used.

Recent advances in 3D acquisition have brought the possibility of capturing high resolution scans of rough surface with great accuracy. In fact, surface details that can be captured by photometric stereo techniques are limited only by sensor resolution. We think that these advances should profit 3D surface texture characterisation.

As surface normals can be represented as a needle-map in a 2D image, one might be tempted to apply classic texture analysis methods directly to these. However the problem is that, unlike pixel-colour, normals do not lie in a linear space (adding two normal vectors does not result in a normal unless further normalisation is performed). Thus linear operations such as those required by most of the 2D texture analysis method should not be applied directly to normals.

The methods we introduce in this work produce illumination/view independent texture descriptors, taking into account the non-linearity of the space where the normals lie and operate directly on the surface 3D properties. This is in contrast to the BTF-related methods, which do not have a knowledge of the underlying surface geometry. The new methods are intuitively closer to the classic texture representation methods, do not necessarily require heavy data compression and have their accuracy conditioned only by the surface normal acquisition quality.

1.3 Contributions

The main contributions of this thesis are covered in Chapters 4 and 5. These include:

- The collection of an extensive dataset of high resolution facial scans with various skin conditions along with perceptual human rating according to these conditions.
- A comparative study of 2D and 3D skin texture characterisation.
- The introduction of three methods for describing 3D surface texture.

We give a brief overview here of each of these contributions. These are covered in more detail in Chapters 4 and 5.

1.3.1 The dataset

We have built a Photometric Lightstage and used it to collect an extensive dataset consisting of very high resolution scans of human faces presenting various skin conditions (Figure 1.2).

The dataset has been annotated through experiments where human participants were asked to rate skin patches on three skin conditions: wrinkles, large pores and acne. The patches are extracted from all the faces according to a template which divides a face into 10 regions (Figure 1.3). We assume that the three skin conditions we are interested in (wrinkles, acne and large pores) mostly occur in

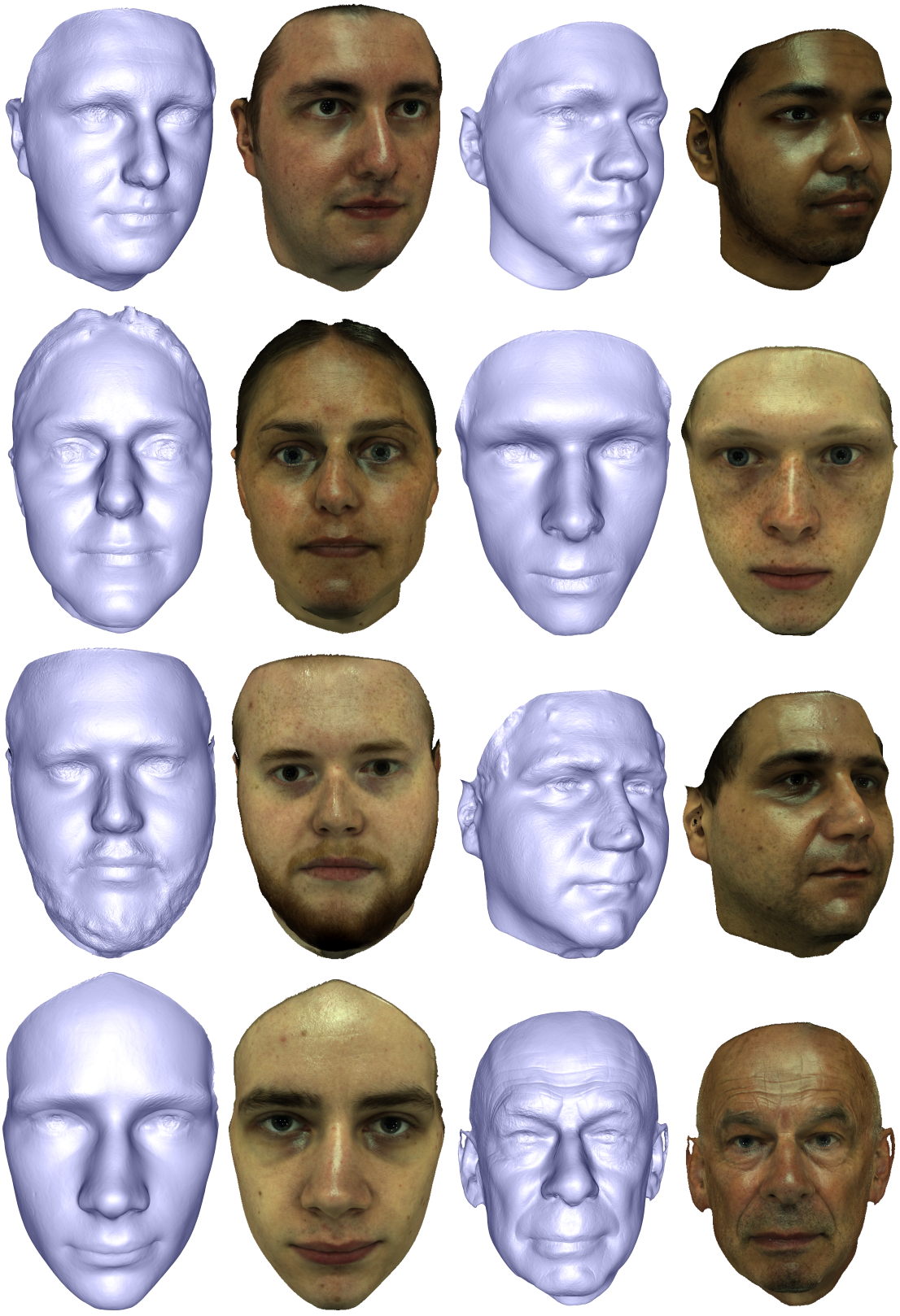


FIGURE 1.2: Data collection (example 3D face scans)

only three of these ten facial regions: forehead, cheek and eyelid/eye-corner. So only these three regions were rated.

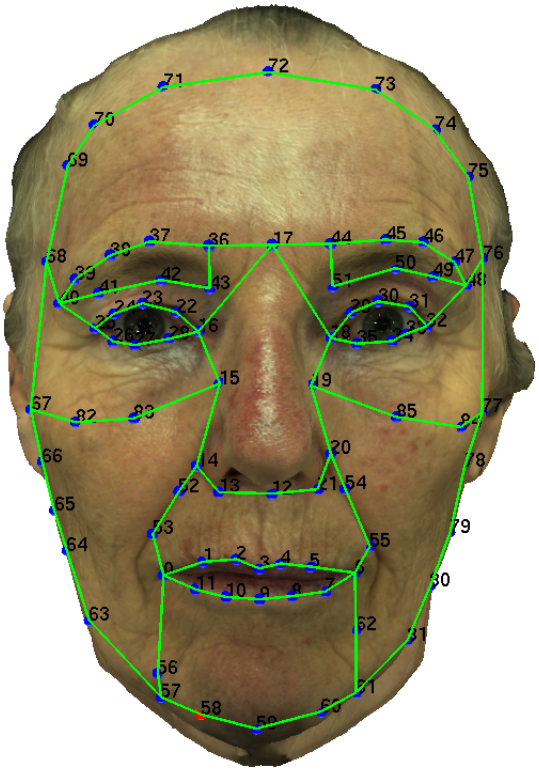


FIGURE 1.3: Facial regions segmentation

In order to get reliable ratings, providing the raters with photo-realistic renderings of the skin patches was critical. We used a Cook-Torrance [9] Bidirectional Reflectance Distribution Function (BRDF) along with a hybrid bump-mapping method which achieved excellent realism on the rendered skin patches, accurately showing the skin fine details. Figure 1.4 shows examples of rendered skin patches.

For each skin condition each patch was rated on a scale going from 1 to 5, 1 meaning "total absence" and 5 meaning "high presence". Figure 1.5 shows an illustration of our rating platform.

We give in Chapter 5 a detailed description of the Lightstage, the collection process and the data processing.

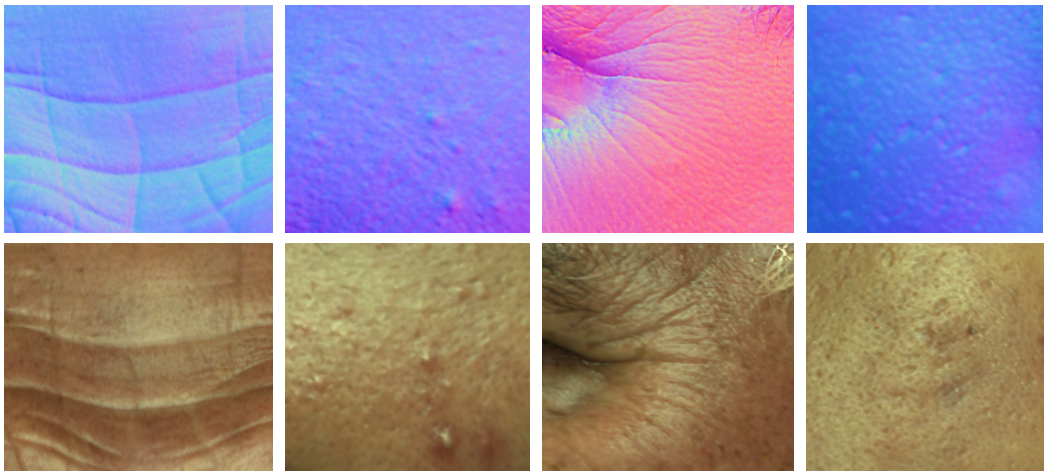


FIGURE 1.4: Surface normal orientation (top row) and their rendering (bottom row) of a few facial regions

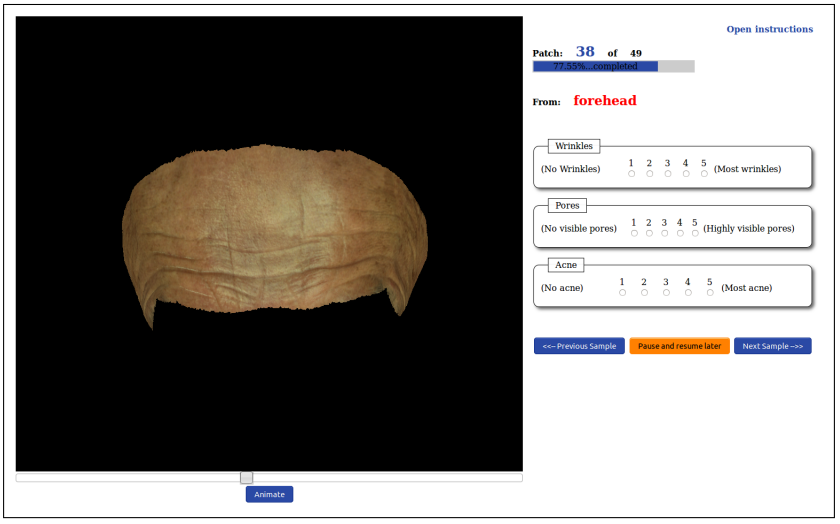


FIGURE 1.5: The skin patch rating platform

1.3.2 Comparative study of 2D vs 3D Skin Texture Analysis

To support our primary motivation of improving texture analysis on rough surfaces – particularly on facial skin – with 3D geometrical features, we conducted a comparative study [75] where classic 2D texture descriptors (LBPs, Gabor Wavelet) are extracted from skin albedo and simplistic 3D extensions of these descriptors are extracted from skin normal maps. We then evaluated both the 2D

and 3D texture descriptors on classifying skin wrinkle and large pores. Results showed a significant improvement when the 3D descriptors are used.

The extension of the 2D texture descriptors to 3D data were based on a rather simple idea. As the normals do not lie in a linear space, we first put the normals in a representation suitable for linear operations, and then apply the classic texture analysis methods on the target space. We have tried two representations. In the first representation, we used the spherical coordinates of the normal and in the second, we computed the geodesic coordinates of the normal by projecting it on the tangent plane.

1.3.3 3D Surface Texture Methods

- **Rotation Fields Pyramid** In this approach, we extend to surface orientation data the notion of image pyramid which is widely used in 2D texture analysis [2, 36, 69]. Famous representations include the Laplacian pyramid and the Steerable pyramid. The main idea is to separate at different scales the high frequency information present in the image from the low frequency. We propose a multi-resolution analysis scheme where at each level of the pyramid, the high frequency information present in the normal map is extracted in the form of rotation fields. We also demonstrate that these Rotation Fields Pyramid can be used to synthesise three dimensional surface texture.

- **Local Orientation Patterns** This method is based on the notion of Texture Spectrum introduced in [85] in the early nineties and which has inspired widely used texture descriptors such as the Local Binary Patterns. The main idea here is to define pattern functions that evaluate how the normals vary at each neighbourhood and plug these into a Texture Spectrum framework. We have proposed two pattern functions. The first function applies a threshold to the angles between the central normal and the surrounding ones. The second function directly compares the normals' azimuthal and elevation angles.

- **Multi-scale APDI** In this work, we extend and improve the APDI (Azimuthal Projection Distance Image) descriptors first proposed in [72]. We first reformulate the projection distance computation which only took into account the polar angle of the normal. We then introduce a multi-scale scheme with a geodesic-based algorithm for scaling the normal maps through the different levels of the analysis pyramid.

1.4 Thesis Outline

The rest of the thesis is organised as follows:

Chapter 2 gives a thorough analysis of a number of related studies. It is split into two parts.

- The first part concerns 3D surface texture analysis. It gives a state of the art in the field by presenting a number of studies on 3D surface texture characterisation expanding from the early works [48] and [12] to the more recent work [72].

- The second part presents a survey on skin imaging methods and automated analysis. A number of works on using computerised systems to capture, analyse or diagnose various skin-related conditions are presented and discussed.

Chapter 3 presents a comparative study between 2D and 3D surface texture analysis.

Chapter 4 describes the novel methods of 3D surface texture analysis we propose. These consist of:

- Simple extension of 2D texture descriptors (LBPs and Gabor Wavelets) to 3D data

- Multi-Scale APDI

- Local Orientation Patterns

- Multi-Scale Geodesic Wavelet

Chapter 5 is dedicated to the ground truth description. It comprises two sections:

- The first section describes the data collection process and the capture device (i.e. the lightstage).

- The second section concerns the data annotation. It gives a detailed description of the annotation process and presents some statistical and perceptual observations on the human ratings.

Chapter 6 presents our experiments and results of classifying wrinkles, large pores and acne using some of the state-of-the-art 3D surface texture characterisation methods and our proposed methods.

Chapter 7 discusses the results, gives some future directions and concludes.

Chapter 2

Literature Review

2.1 Introduction

As our main objective is to exploit facial skin textural information from high resolution 3D data to assess the presence of certain skin conditions, this thesis covers two principal issues:

- 3D surface texture analysis
- Skin imaging and automated analysis

In this chapter, I present various studies related to these two topics. However, as a number of 3D surface texture analysis methods presented here are based on extensions of 2D texture analysis methods, we first give a rapid overview of the existing 2D methods (a more exhaustive survey of 2D texture measures can be found in [96]).

2.2 2D Texture Analysis

Texture is ubiquitous in the perception of natural surfaces and its digital representation is fundamental to Computer Vision. Texture characterisation is key to a number of visual computing related applications such as object recognition, content-based image retrieval and computer graphics. A number of efficient and powerful 2D texture analysis methods have been proposed in the literature. These methods can be divided into three categories:

2.2.1 Statistical methods

These methods assume that the texture is fully determined by the spatial distribution of pixel values in the image. One of the earliest example is the use of the Grey Level Co-occurrence Matrix (GLCM). Haralick proposed a set of measures (uniformity, entropy, contrast, inverse moment, correlation, etc.) based on the GLCM to describe different textural properties [35]. The GLCM is a matrix that represents the co-occurrence distribution of pixel values within a given window in a grey level image. It is a square matrix whose size is determined by the pixel depth. For example a grey level image with 8-bit pixel will yield a $2^8 \times 2^8$ GLCM. Given a $w \times h$ grey level image and a $u \times v$ window, the GLCM is built using the following formula:

$$C_{u,v}(i,j) = \sum_{p=1}^{w} \sum_{q=1}^{h} \begin{cases} 1 & \text{if } I(p,q) = i \text{ and } I(p+u, q+u) = j \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

Another widely used statistical approach for characterising texture is the autocorrelation function. It is mostly used to detect repetitive texture patterns. It consists in a normalized scalar product between an image and its shifted versions. If I is an image of size $w \times h$, the autocorrelation function $f(k,l)$, with shifting parameters k and l, is computed as:

$$f(k,l) = \frac{\frac{1}{(w-k)(h-l)} \sum_{i=1}^{w-k} \sum_{j=1}^{h-l} I(i,j)I(i+k,j+l)}{\frac{1}{w \times h} I^2(i,j)} \tag{2.2}$$

As well as these global approaches, there are some powerful local methods where the statistics are computed in the immediate neighbourhood of a particular pixel. These include various descriptors proposed in [67], namely the Symmetric Autocorrelation function (SAC) and its extensions (SRAC and SCOV) and the famous Local Binary Patterns (LBPs) which are one of the most used texture descriptors in the literature. We present these in more details in Section 3.3.1. A number of variants of the LBPs are also proposed in the literature. [42] proposes the transition Local Binary Patterns (tLBP) and the direction coded Local Binary Pattern (dLBP). In the first variant the patterns are obtained from comparing the neighbouring pixel to each other instead of the central pixel and in the second variant the variation direction between central pixel and neighbour is also encoded. [97] introduces Multi-block Local Binary Patterns (MB-LBP) where, instead of comparing individual pixels, the authors compare mean pixel value of fixed-size blocks which has shown improvement in face detection task. [76] proposed the Advanced Local Binary Patterns (ALBP) which, as with the extension already proposed in [63], introduce rotation-invariant encodings.

2.2.2 Structural methods

These consider texture as a structured layout of texture primitives also called texture elements. Such methods divide into **geometrical** and **topological** approaches. In geometrical approaches, coarse geometrical properties such as perimeter and compactness are used to characterise texture primitives [29]. Topological approaches use various filtering methods to extract primitives such as lines, edges, blobs, etc. The texture descriptor is then made of different properties of these extracted primitives, namely, number, orientation, density, etc. [21] investigates extraction of regular shape such as lines using a Hough Transform [18].

2.2.3 Model-based methods

In these methods, texture is represented with either a probabilistic model or a projective decomposition along a set of basis vectors. These representations require the determination of a certain number of parameters or coefficients to characterise the texture. The **Markov model**-based methods constitute an important subset of these methods. Hidden Markov Models (HMMs) have been extensively used to characterise texture [40, 89]. [8] used a Gaussian Markov Random Field (GMRF) to model rotated and scaled texture. Methods using **sub-band decomposition** techniques include the wavelet transform [55, 68], the steerable pyramid [31] and the Gabor Bank of filters [26, 39, 68].

2.2.4 Summary

We have presented in this section three categories of 2D Texture analysis methods. Table 2.1 summarise these along with their strengths and weaknesses. Generally, the approach chosen for a specific application depends upon the aspect of texture one wishes to capture. Most of these 2D texture analysis methods have proven their efficiency through various applications. However, all 2D methods make the assumption that apparent texture is independent to illumination and viewpoint, which make them not robust to imaging condition change as stated in the Table 2.1. While the assumption can be approximated when studying smooth surfaces, the apparent texture of surfaces involving rough relief is considerably affected by illumination and viewpoint.

2.3 3D Surface Texture Analysis

2.3.1 Introduction

As stated in the previous section, texture analysis has historically been considered to be a 2D problem in digital image analysis and computer vision. This has lead to neglect of the effect of imaging conditions such as illumination and viewpoint on the apparent texture of rough surfaces. Generally the appearance of natural surfaces is not only determined by their intrinsic reflectance properties (color or albedo), but is also considerably affected by the interaction between their geometrical structure, the light and viewpoint.

Methods	Description	Strength	Limitations
Statistical Methods	Modelling spatial distribution of the pixel in the image and their relations	Versatile as prior knowledge of structural organisation of texture is not needed	Not robust to lighting condition change
Structural Methods	Extracting regular texture primitives or "texels" and modelling their arrangement	Good symbolic representation of texture. More intuitive than statistical methods	"Texel" might be hard to extract from natural/stochastic texture. Not robust to lighting condition change
Model-based Methods	Representing texture through a parametric model	Generative and useful for texture synthesis	Generally expensive computations for determining parameters of the model. Not robust to lighting condition change

TABLE 2.1: Summary of existing 2D texture characterisation methods

Around 20 years ago, a new trend aimed at addressing these issues emerged. These new studies all take into account the effects that illumination and viewpoint have on apparent textures. In the rest of this thesis, we will refer to these types of texture method, sensitive to illumination/view changes, as **3D surface texture**. In this section, we review these 3D surface texture analysis methods. These can be categorised into three families:

- 3D texton based methods
- Bidirectional texture based methods

- Geometrical methods

2.3.2 3D texton based methods

The notion of the 3D Texton was introduced in [48] and has been widely used/extended since to represent natural surfaces' visual appearance [54, 57, 61, 66, 92]. The main idea is to simultaneously encode the two attributes that most affect how a surface is visually perceived; these are the surface normals and reflectance properties. To characterise a given surface's texture, the approach consists of exploiting filter responses on several images of the same surface taken in different imaging conditions (illumination/viewpoint). In addition, from the observation that most natural textures present spatial redundancy, the authors proposed to quantise these filter responses into a reduced set of texture prototypes. This results in a dictionary of tiny texture patch representations called *3D textons* that covers all possible local surface configurations.

In the original formulation [48], given a set of images of various surfaces at different light directions and viewpoints, a filter bank of 36 oriented, scaled and phased filters is applied to each image. If $N_{v,l}$ is the number of viewpoints and light directions, this produces for each sample $36 \times N_{v,l}$ filter responses which are all concatenated and clustered using K-means. Each resulting cluster corresponds to a 3D texton. The Columbia-Utrecht dataset [12] (60 different materials, each with 205 images of different illumination/viewpoint) is used to construct a universal dictionary of 100 3D textons (only 20 images randomly selected for each material in [12]). The choice of how many clusters is critical

as it determines the expressiveness of the vocabulary. Though [48] fixed their cluster count at 100 (preceded by an initialisation where 400 clusters are used before being merged down to 100), [84] argued that the use of extremely reduced number of clusters number along with a rotation-invariant filter bank improves classification results.

2.3.3 Bidirectional texture based methods

In contrast to the 3D texton-based methods, the Bidirectional Texture Function (BTF) operates at a higher level of abstraction for representing surface properties that affect the apparent texture. This makes them useful for analysing as well as for synthesising natural texture. However, when used for analysis, they are generally combined with a texton-based quantisation layer. The notion of Bidirectional Texture Function was first introduced in [12] and constitutes the most advanced and accurate representation of natural surfaces visual properties to date [22].

As the Bidirectional Reflectance Distribution Function (BRDF) models a surface's reflectance as a function of illumination and viewpoint, the BTF models a surface's texture as a function of illumination and viewpoint. It is a seven dimensional function and represents texture as a function of the spectral band, the planar position, the view and light directions:

$$BTF(r_x, r_y, \rho, \phi_i, \theta_i, \phi_v, \theta_v) \tag{2.3}$$

where r_x and r_y are the horizontal and vertical positions respectively, ρ is the spectral band, ϕ_i and θ_i are the elevation and azimuthal angles of the light direction respectively and ϕ_v and θ_v the elevation and azimuthal angles of the viewing direction respectively.

BTF measurement generally involves a very complex setup where automated devices allow coordinated changes of either the lighting conditions or the camera viewpoint or, in some systems, both. A considerable number of BTF measurement methods (along with rich datasets) have been proposed in the past [12, 33, 41, 56]. While they differ in the type of setup used (fixed light and moving camera, fixed viewpoint and moving light, or moving camera and light) they all result in a big database of samples of different materials under various lighting and viewing conditions.

Although BTF is extensively used in Computer Graphics, generally for photo-realistic texture synthesis and rendering purposes, it is also used to create and evaluate texture features that are robust to imaging conditions (illumination/viewpoint). [11] analysed skin texture using a BTF made of more than 3500 images to discriminate between skin disorders like acne and psoriasis. They proposed two models for representing the skin texture:

- **Image Textons**: where a bank of filters is applied to the BTF database, followed by clustering the filter responses to a library of *image textons*. These *image textons* differ from the *3D textons* [48] in that the feature vectors composing the input of the clustering are not aggregated through

change of illumination/viewpoint. Instead, for each sample, the filters responses at a pixel constitute a full feature vector [10, 11]. This makes the registration step required in the *3D textons* approach [48] unnecessary.

- **Symbolic Texture Primitives**: where, at each pixel, the index of the filter from the bank that gives the maximal response is symbolically assigned to that pixel. Then the texture primitives are built by locally concatenating the symbolic labels of pixels within an arbitrary region.

[80] introduced the notion of **dimensionality surface** as a measure of appearance variability due to the effects of viewpoint and illumination changes on fine surface geometry. From the CURet Bidirectional Texture database [12], they applied a set of multi-band correlation functions $R_{ij(m,n)}$ on each image of each material sample (i and j being spectral bands and $[m, n]$ an image region). For an RGB image, this yields an inter-band correlation vector with a length of $9L$ (with $L = (2m + 1)(n + 1)$). If C denotes the number of images with different illumination/viewpoint for each material sample, a matrix $R_{9L\times C}$ is obtained. Finally the **dimensionality surface** is defined as the rank of the smallest orthogonal basis of $R_{9L\times C}$. These basis vectors represent 3D texture primitives for each material sample over all the illuminations and viewpoints.

[7] introduced the KTH-TIPS2 material database (11 materials each with 4 different imaging conditions) and used it to test the robustness of various state-of-the-art texture descriptors to pose and illumination changes. They experimented with successively including in the training set samples from 1, 2 and 3 pose/illumination conditions and testing with samples from the unseen pose/illumination

conditions. One of their findings was that the more sample groups they add to the training set the better the classification method performs.

More recent studies include [49] which proposes learning discriminative models for determining optimal texture filters for given illumination conditions. The BTF database is collected using a dome of controllable LEDs and a fixed camera on the top. The acquired database consist of 90 material samples (of aluminium, stainless-steel, granite, carpet, paper, plastic and wood) captured under 6×25 lighting patterns (6 spectral bands and 25 lighting directions). To learn optimal filters as a function of lighting conditions, they express the filter responses on 3D-textured surfaces under a given lighting direction over the neighbourhood N as follow:

$$r = \sum_{p \in N} \alpha_p I(p) = \sum_{p \in N} \sum_{j \in M} \alpha_p \cdot f(p, j) \cdot L_j \tag{2.4}$$

where α_p is the filter weight at pixel p; $I(p)$ the observed intensity at pixel p and is given by the dot product between the apparent BTF $f(p, j)$ and the incident illumination L_j, $j = 1, ..., M$. For a bank of K filters, they obtain a K-length vector:

$$r = [r^1, ..., r^k]^T = (A^T \cdot F^T) \cdot L \tag{2.5}$$

(with A the $N \times K$ matrix of filter weights and F the $M \times N$ matrix of apparent BTFs). The objective is to determine the filter weights given a known incident

illumination, and the authors model each filter in A as a linear composition of Gabor filters, S-filter (introduced in [74]) and maximal response filter. If the set of basis filters is denoted by the matrix B and the weights attributed to each filter by the matrix W, the filter matrix $A = BW$ and the feature vector becomes:

$$r = W^T \cdot B^T \cdot F^T \cdot L \tag{2.6}$$

To find the joint optimal filter coefficients and illumination pattern, W and L are optimised alternately using the BTF database (where B^T and F^T are known).

Finally, [88] recently proposed to bypass the limitation of a fixed number of illumination patterns in most of the BTF databases, by generating synthetic training samples with arbitrary illumination patterns from a measured BTF. They acquired a BTF of 7 materials, each of these imaged under 12 different lighting conditions. From this BTF, they synthesised new samples with arbitrary lighting conditions and used these training samples. To encode the surface texture they used a clustering scheme on SIFT extracted features to build a dictionary of texture primitives.

2.3.4 Geometrical methods

The methods presented in the two preceding sections are image-based as the intrinsic geometry of the material's surface is not known. The considerable number of image samples (different illumination/viewpoint) needed by these image-based

methods in order to capture the three-dimensional properties of the studied surfaces makes their use excessively demanding in storage capacity and generally leads to very long feature vectors.

Some recent works have looked at characterising the 3D texture of surfaces directly from their measured fine geometry, which provides a more compact representation of the intrinsic three-dimensional properties. These studies leverage the considerable advances made over the last decades in 3D surface capture.

[77] proposed to compute a co-occurrence matrix from the orientation of measured surface normals. Their method involves quantising the normals' orientation into a discrete space. For each normal, the slant and tilt angles are discretised in three equal intervals. This results in 9 levels upon which the co-occurrence matrix is constructed.

[72] extracted Local Binary Pattern features from two different 2D representations of 3D geometrical data to classify 3D facial action units. The two representations are:

- **depth map**, which is a very simple two-dimensional representation of a three-dimensional surface. It encodes in a regular grid of pixels the depth of each point of the 3D surface (its distance to the image plane) and generally involves interpolating between the points to get a dense map.

- **azimuthal projection distance image (APDI)** , which encodes the surface orientation in a 2D image. The idea is to project each surface normal onto the tangent plane and take the L2-norm of the projected point

as a grey level. For a grid of surface normals, the azimuthal projection of a normal at the node (i, j) is given by:

$$x_{i,j} = k' \cos\theta_{i,j} \sin\left(\phi_{i,j} - \bar{\phi}_{i,j}\right) \tag{2.7}$$

$$y_{i,j} = k' \left(\cos\bar{\theta}_{i,j} \sin\phi_{i,j} - \sin\bar{\theta}_{i,j} \cos\theta_{i,j} \cos\left(\phi_{i,j} - \bar{\phi}_{i,j}\right)\right) \tag{2.8}$$

where $\theta_{i,j}$ and $\phi_{i,j}$ are the polar and azimuthal angles of the surface normal respectively. $\bar{\theta}_{i,j}$ and $\bar{\phi}_{i,j}$ are the polar and azimuthal angles of the mean surface normal over a fixed neighbourhood around (i, j) respectively. Finally $k' = \frac{c}{sin(c)}$ with $c = \sin\bar{\theta}_{i,j} \sin\theta_{i,j} + \cos\theta_{i,j} \cos\left(\phi_{i,j} - \bar{\phi}_{i,j}\right)$

To compute the APDI, a constant mean surface normal $(0, 0, 1)$ (z-axis direction) is fixed which leads to $\bar{\theta} = \frac{pi}{2}$, $\bar{\phi} = 0$ and $c = \sin\theta_{i,j}$. Thus Equation 2.7 and 2.8 become:

$$x_{i,j} = k' \cos\theta_{i,j} \sin\phi_{i,j} \tag{2.9}$$

$$y_{i,j} = k' \cos\theta_{i,j} \cos\phi_{i,j} \tag{2.10}$$

Each pixel value of the Azimuthal Projection Distance Image (APDI) is given by the $L2$-norm of $(x_{i,j}, y_{i,j})$:

$$APDI_{i,j} = \sqrt{x_{i,j}^2 + y_{i,j}^2} \tag{2.11}$$

The APDI constitute one of the most recent and advance geometrical 3D surface texture representation to our knowledge. However they were design to extract coarse 3D deformation such as facial expression. One of the novel methods proposed in this thesis extends these in order to make them applicable to fine texture characterisation such as the skin conditions we are addressing in this study.

2.3.5 Summary

In this section, we have reviewed several approaches for characterising view and illumination-dependent texture, which we refer to in this work as **3D surface texture**. We have shown that the existing methods in the literature can be divided into three categories: 3D texton-based, BTF-based and geometrical. The first two classes are image-based as the intrinsic geometry of the surface is not known. They extract texture primitives from several samples of the same surface imaged under different conditions (view/illumination). The last category includes methods that characterise texture from the surface's measured geometry. These constitute a nascent trend and aim to take advantage of the considerable advances realised in 3D surface capture. As such these methods provide surface texture representations more robust to imaging conditions than the classical two-dimensional methods. Table 2.2 gives a summary of these methods along with their strength and limitations.

Methods	Description	Strength	Limitations
3D texton [48, 54, 57, 61, 66, 92]	Vocabulary of small surface patches with corresponding 3D properties	Samples aligned prior to clustering yielding more appropriate representation of natural surfaces. Can also be used for light/view prediction	Require image registration and also extensive dataset. Produce large feature vectors
BTF methods [12, 33, 41, 56]	Database of surface texture variations due to viewpoint and illumination changes	Easier to generate as no image registration is required. Can be also used for surface synthesis.	Less robust to cast shadow and occlusion as result of non image registration. Also require extensive dataset and produce large feature vectors
Geometrical Methods [72, 77]	Texture features extracted from the surface fine 3D geometry	Yield more compact representations. Operate directly on the surface geometry and thus more robust to cast shadow and occlusion. Benefit from recent advances in 3D surface capture techniques	Nascent trend with a limited number of works to date. Require high resolution 3D data.

TABLE 2.2: Summary of existing 3D surface texture characterisation methods

2.4 Automated Skin Capture and Analysis

2.4.1 Introduction

While the design of novel 3D surface texture descriptors constitutes the principal motivation of this thesis, our first objective is to use these descriptors for analysing human facial skin in order to detect conditions such as acne, wrinkles and visible pores. Thus our project includes the imaging of three dimensional fine structure of human skin and their analysis using the 3D surface texture characterisation methods developed in this work.

In this section, we cover the state-of-the-art in 3D imaging of human skin as well as automated skin condition detection and/or classification.

2.4.2 3D skin micro-structure imaging

Human skin micro-structure imaging has seen considerable interest over the last decade. This interest has been motivated by medical (dermatology) applications and the increasing demand for photo-realistic solutions from the game and film industry.

[11] used a bidirectional imaging system to capture the micro-structure of skin regions affected by diverse dermatological disorders (psoriasis, acne, contact dermatitis, etc.). They collected a publicly available dataset of more than 3500

images: the **Rutgers Skin Texture Database**. They used two different mechanical setups that allowed them to capture skin regions in various viewpoints and light directions.

[53] proposed a spherical gradient illumination-based system to capture three-dimensional facial skin structure down to the level of the pores. The spherical gradient illumination environment is obtained using a dome of controlled light called the Lightstage. They combined this with a polarised light technique to separately capture the diffuse and specular surface properties. The resulting data are in form of normal maps. They have shown that the specular normal maps capture most of the surface details while the diffuse ones are more subject to sub-surface scattering. These polarisation and wavelength-dependant measurements constitute useful data for understanding how the human skin interacts with light as well modelling its micro-structure.

Many improvements and applications have been added to the capture system since. [30] proposed a measurement-based synthesis of facial micro-geometry where they measure the micro-structure of the skin patches using a twelve-light hemisphere being able to emit cross-polarised light. The acquired skin micro-structure is used to extract displacement maps which is then synthesised using an image analogy method [37] which consists in learning transformations from paired image data and apply these to new images. [59] extended this to dynamic skin micro-structure deformation.

Another skin reflectance measurement using a Lightstage is conducted in [90]. They augment their data with an extra skin subsurface scan using a fiber optic

spectrometer which is a more sophisticated device allowing measurements of sub-surface properties such as haemoglobin or glucose concentrations. The authors also fitted the analytic BSSRDF (Bidirectional Sub-Surface Reflectance Distribution Function) proposed in [34] to their measured data and conducted analysis on the relations between the BSSRDF parameters (scattering and absorption coefficients) and various attributes of the subject such as age, skin type, etc.

PRIMO [1] is a commercial solution for 3D skin measurements used in some automated skin disruption detection studies such [44]. It is a hand-held optical-based system using structured light and a high resolution censor allowing measurements of skin micro-topography and roughness with a field of view of $45 \times 30 \times 30mm$

2.4.3 Automated skin disruption analysis

Automated detection of human skin disorders has aroused considerable interest in the computer science and image analysis communities. Many studies have proposed automated systems for detecting and/or classifying various skin disorders. While a great number of these studies were interested in the classification of benign and malignant melanoma, others addressed more common skin conditions such as acne, contact dermatitis, wrinkles etc.

[12] employed image-based models learnt on a BTF database to classify various skin disorders such as acne, congenital melanocytic nevus, keratosis pilaris, psoriasis, contact dermatitis and other conditions. The BTF database, called the Rutgers Skin Texture Database, contains over 3500 images of normal and

[1] http://www.gfm3d.com/

affected skin. Two models of extracting texture primitives from the BTF images are proposed. The first is based on an *image texton* approach and the second, on a *symbolic texture* representation.

[25] used skin line patterns extracted from simply captured white light optical skin images to classify benign moles and malignant melanoma. The skin line patterns were extracted by high-pass filtering and enhanced by adaptive anisotropic filtering. The skin line direction variation was also estimated using a local image gradient matrix. Then the difference between these two measures was used as a feature. ABCD features (Asymmetry, Border irregularity, Colour and Diameter) have also been used by the same authors for classifying skin melanoma [23, 24]. The ABCD features were introduced in [45], and consist of a set of values computed using morphological analysis on the affected region. They characterise the size, shape and appearance of the lesion, and have been shown to be considerably discriminant in benign and malignant melanoma classification.

[16] have investigated augmenting these ABCD features with 3D texture primitives and have found that the extra information enhances classification results. This can be explained by the fact that melanoma is not a flat two-dimensional object, but a three-dimensional one. The approach consisted of analysing the surface changes in the tilt and slant direction of skin normals. They simulate healthy skin regions with a two-dimensional Gaussian function and then compare these with the observed tilt and slant patterns of lesions. The difference of the two patterns is considered as a characterisation of the perturbation and is used to classify the two types of lesion.

Another study on melanoma classification using 3D features has been conducted in [86]. Here differential forms, i.e., first- and second-order partial derivatives of the 3D skin surface are used to describe the lesions. The authors first used photometric stereo techniques to calculate the surface normals, and then separated the lesion areas from the unaffected skin using the principal curvatures and the second fundamental form computed from the measured surface normals.

[44] used a commercial 3D skin imaging system (PRIMO [2]) to quantify skin surface roughness and acne volumes in order to evaluate the efficiency of anti-acne treatment. The skin surface roughness was measured using four parameters: the arithmetic average of the absolute difference between point heights and the average surface height; the arithmetic average of peak-to-valley values of five consecutive sampling sections; maximum peak-to-valley value; and maximum peak height. Acne lesions were represented by their total number and their average volume in a specified sector.

2.4.4 Summary

In this section the state-of-the-art in human skin micro-structure imaging techniques and automated skin disorder detection/classification were discussed. It is clear that enormous advances have been made in imaging techniques, as it is now possible to capture the skin's three-dimensional micro-structure down to the level of the pores. However, it seems that these newly available possibilities

[2] http://www.gfm3d.com/

for data capture are not fully exploited on the analysis side as most of the studies presented above use either two-dimensional image-based texture features or rather coarse three-dimensional surface properties. One of the few studies that exploited the skin three-dimensional micro-structure used a BTF representation [12] which, even though taking into account changes in illumination/viewpoint, is still an image-based representation as the underlying surface geometry is not known.

2.5 Conclusion

This chapter presented the state-of-the art in 3D surface texture analysis and skin micro-structure imaging techniques. We have seen that the existing 3D surface texture characterisation methods can be divided in 3 categories:

The 3D texton methods aim at creating vocabularies of surface patches along with their corresponding 3D properties. The vocabulary is then used as a basis to represent the intrinsic 3D texture of the surface. Before generating the vocabularies, an image registration is performed on the dataset between samples of the same surface. While costing an extra computation stage, it makes the approach more robust to shadow and occlusion than the BTF based methods.

The BTF based methods extract the 3D texture descriptors from a BTF database which represents a large collection of texture samples under various lighting and viewing conditions. They do not require registering the samples captured across viewpoint and light change which make them computationally less demanding

than the 3D texton methods but at the same time, as mentioned earlier, less robust to shadow and occlusion.

Both the 3D texton and the BTF based methods tend to use extensive dataset and produce large feature vectors. This makes the geometrical methods, which operate directly on the surface geometry (more compact representation than BTF data), more attractive. Furthermore, the geometrical methods can take advantage of the recent considerable progresses made in 3D surface fine structure capture.

However, the geometrical methods are at an early stage of development. To date there only a few studies that address the issues of 3D surface texture with a geometrical approaches. The most popular studies include [72, 77].

In Chapter 5, we propose three novel methods of characterising 3D surface texture from geometrical data. One of these method will be an extension and improvement of [72]. Their performances will be evaluated on classifying skin conditions such as wrinkle, acne and pores using our collected dataset of high resolution 3D facial scans.

Before introducing these novel methods, we present in the next chapter a comparative study aiming at demonstrating the advantage of considering texture as 3D problem rather than a 2D one. The experiment consists in comparing skin condition classification performances using classical 2D texture features (Gabor and LBPs) and simplistic extensions of these to 3D data.

Chapter 3

2D Texture vs 3D Skin Texture Analysis

3.1 Introduction

In this chapter, we demonstrate the added value of using 3D geometric features for surface texture analysis. For this, we compare techniques for skin texture classification using various 2D texture descriptors and their extensions to 3D surface orientation data. We perform a multi-resolution analysis on both the 2D and 3D data. Rotation- Invariant Local Binary Patterns and Multiple Orientation Gabor Filters are used to extract 2D texture features from high resolution facial skin albedo patches. In 3D we extract texture features directly from the corresponding normal map patches by proposing extensions of these texture measures in both the slant/tilt and geodesic tangent spaces. We compare the results

on the task of classifying facial wrinkles, acne and pores using both 2D and 3D texture features.

3.2 Motivations

A 2D image is a snapshot of a 3D scene with a set of well-defined conditions, including the lighting environment, the relative position of the camera and the objects forming the scene. These conditions together determine what is seen in the image and what is not. Thus a 2D image will not convey all the information present in the scene it represents and this is also true for the textures present in the scene. It is clear that some texture patterns can be totally captured and rendered in a 2D image without the above conditions considerably altering their appearance: a zebra skin macro-texture won't change considerably in images with changes of lighting condition and viewpoint (as long as there is enough ambient lighting to see the stripes). But when it comes to textures with surface roughness, such as an aerial image of a mountainous region or a bumpy woolen fabric, these imaging conditions can considerably affect their appearance in an image.

In this work we are interested in skin texture, and specifically, variations in skin appearance due to conditions like acne or wrinkles. Despite the apparent global smoothness of its macro-structure, skin texture can be relatively rough in the meso and micro scales. Certain skin conditions such as wrinkles, lines, large pores and moles can cause clearly visible surface irregularities on the skin.

Therefore, skin texture does not escape the issues stated above. We show in Figure 3.1 how the apparent texture of two wrinkly skin patches can change with lighting condition and/or angle of view. Figure 3.1 demonstrates on two patches how the light direction or viewpoint affects the "apparent" amount of wrinkling that is visible.

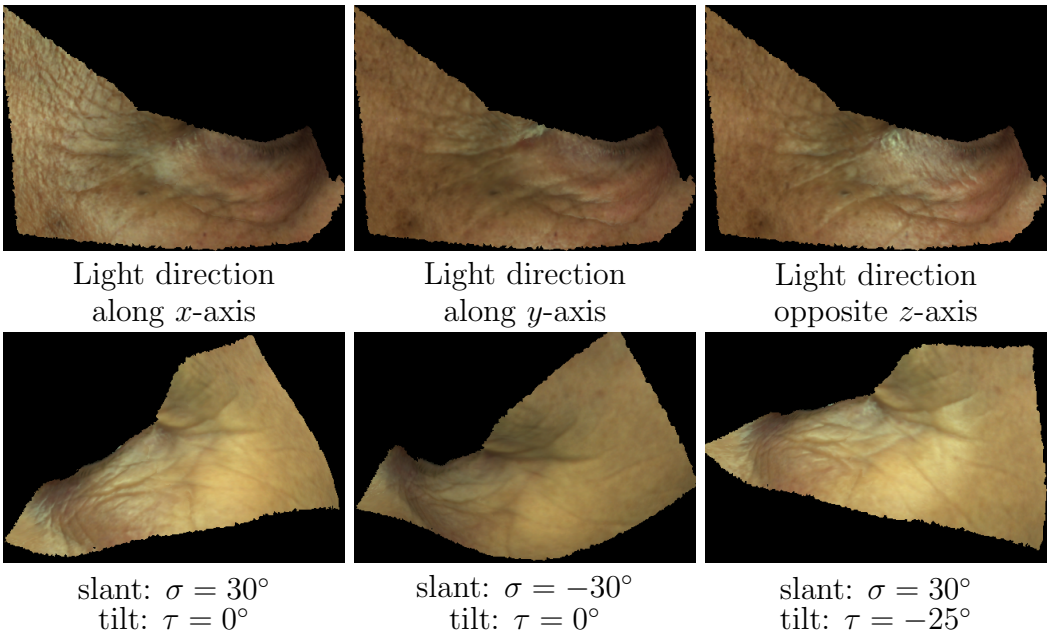


FIGURE 3.1: Change in apparent texture with changes of illumination(top row) and/or viewpoint(bottom row)

It is clear that any 2D texture descriptor would give a different response to each of the images in either the top or bottom row of Figure 3.1, yet they represent the same skin patch on each row. In addition, even if we assume that an image has been captured in optimal position and lighting conditions for best texture and roughness visibility, its 3D representation with the same resolution (which can be obtained with photometric stereo) will always hold more information about the surface texture. This is inherent to the process of recovering a 3D surface which involves change of lighting and/or viewing conditions. To illustrate this,

Figure 3.2 shows the difference between applying a 2D Gabor filter to a skin patch specular albedo (this being the 2D representation with the most surface detail) and on the corresponding normal map (with same resolution as the specular albedo). One can notice that after Gabor filtering of the specular albedo, how the response is mostly confined on the central area, which must be the area with optimal roughness visibility according to the imaging conditions (view/illumination). On the other hand the Gabor response on the normal map is spread all over the patch, suggesting a less view/illumination-dependant response. The filter applied to the albedo is the same than the one applied to the normal map, the only difference is that in the normal map case, it is applied to both the slant and the tilt. The next section gives more details on how we apply filters on the normal maps. Also one can notice how the Gabor response of the normal map seems to show more information than the normal map itself. This is inherent to the way the normal maps are visualised. To be able to display them as RGB images, they are scaled from the $[-1, 1]$ space to the $[0 - 255]$ space which inevitably affects visibility of the intrinsic information. But this is only relevant to visualising the normal map, all the descriptors presented in this work are calculated without scaling.

These observations give a first visual appreciation of the difference between extracting the skin texture from the 2D and the 3D data. In the next section, we conduct a quantitative comparison by classifying skin texture using various 2D texture descriptors and their extensions to 3D data.

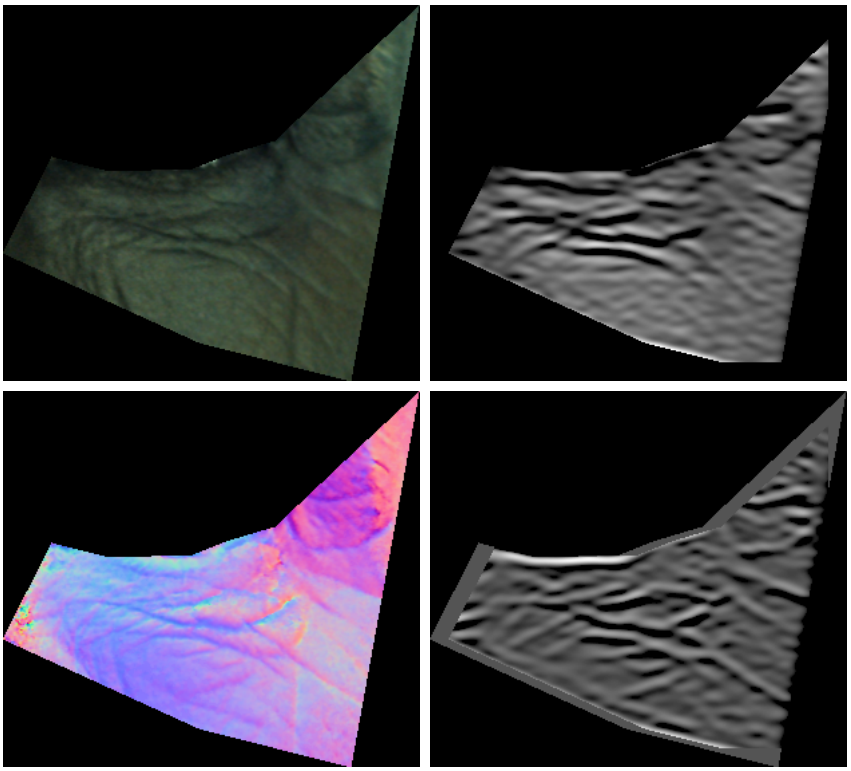


FIGURE 3.2: Gabor responses from the specular 2D image (top row) and the 3D normal map (bottom row). The Gabor filter is set with an orientation of $\frac{\pi}{2}$ a standard deviation of $10 pixels$ and a wavelength of $8 pixels$. The patches' spatial resolution is 310×280

3.3 Skin Texture Classification: 2D vs 3D

In this section, we propose very simple extensions of various 2D texture descriptors and evaluate their performance on classifying the skin texture in comparison with the corresponding 2D counterpart. But, first, we will review the different 2D texture descriptors used.

3.3.1 2D Texture Descriptors

We use multi-orientation and rotation invariant descriptors as most skin condition textures do not have a predominant orientation: a wrinkle is still a wrinkle,

whether it appears vertically, obliquely or horizontally on the face. We describe in this section two 2D texture descriptors which are multi-orientation Gabor filter and rotation-invariant Local Binary Patterns.

Gabor Filtering Gabor filtering is widely used in texture segmentation [26, 39]. It has the advantage of approximating the two-dimensional receptive-field profiles of simple cells in the mammalian visual cortex [13]. The real part of a 2D Gabor filter is represented by the function:

$$f(x, y, \sigma, \theta, \lambda, \gamma, \phi) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \phi\right)$$

which is a sinusoid enveloped in a Gaussian where $x' = x\cos\theta + ysin\theta$ and $y' = y\cos\theta - x\sin\theta$. σ, θ, λ, γ and ϕ are respectively the Gaussian standard deviation, the filter orientation, the bandwidth, the aspect-ratio between the two axes and the phase.

In order to generate multi-orientation texture descriptors, a bank of Gabor filters has to be applied with different values of orientations. It is theoretically not possible to generate perfectly rotation-invariant descriptors, but the more the bank of filters covers varied orientations, the more the resulting descriptors are rotation robust. Figure 3.3 shows the Gabor responses with different orientations on a wrinkly bottom eyelid patch.

Rotation-Invariant Local Binary Patterns Local Binary Patterns were proposed in [62] as an extension of Texture Spectrum, first introduced in the

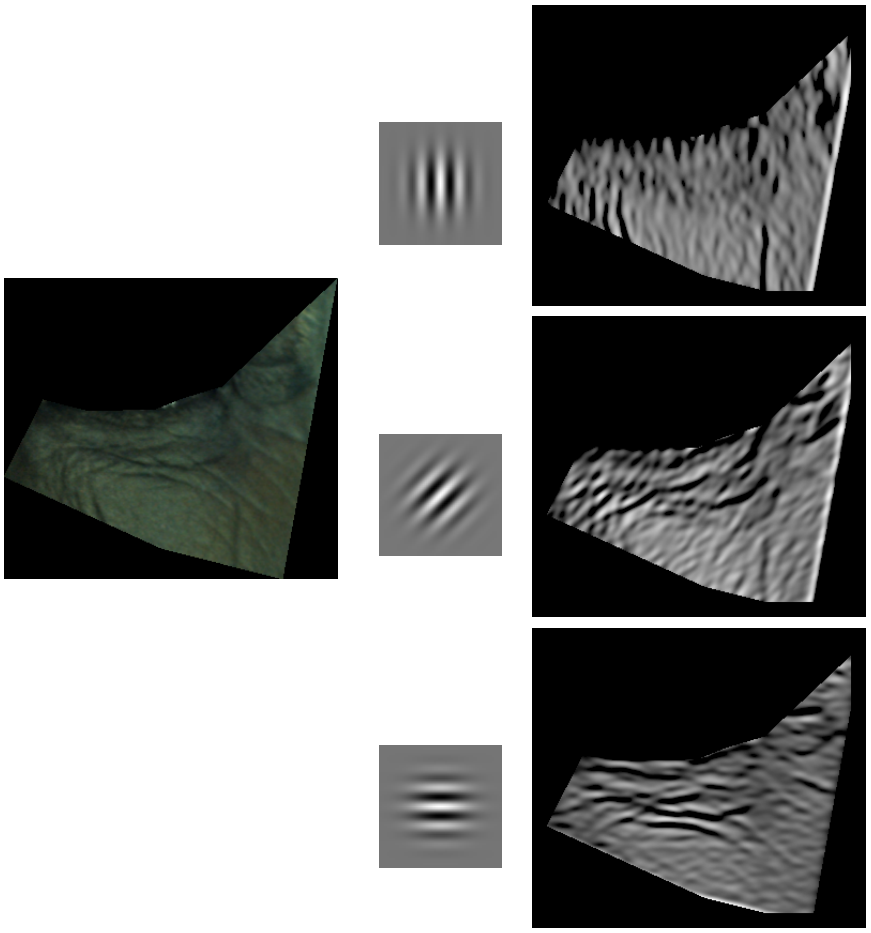


FIGURE 3.3: Gabor responses on specular albedo for different orientations (respectively 0, $\frac{\pi}{3}$ and $\frac{\pi}{2}$) of a wrinkly skin patch surface. The Gabor filter is set with a standard deviation of $6pixels$ and a wavelength of $7pixels$. The images are 310×280)

early nineties [85]. They are one of the most used texture measures in the literature. They consist in thresholding, at any position, the neighbourhood to the central pixel and by multiplying the results by the corresponding binomial weight (Figure 3.4).

In its original formulation, the Local Binary Patterns are not rotation-invariant. [67] proposed a method of getting rotation-invariant Local Binary Patterns. It consists of pre-computing a rotation-free dictionary of patterns, performing a circular shifting on each neighbourhood thresholding result till it matches one in

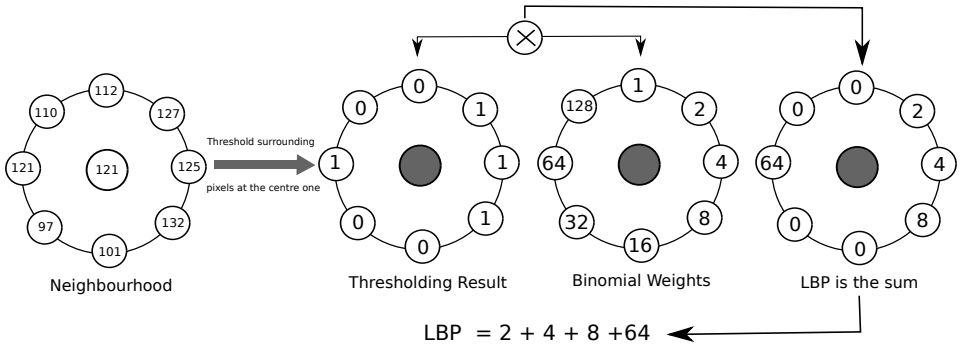


FIGURE 3.4: Local Binary Patterns Chart for an 8-pixels Neighbourhood

the precomputed dictionary and finally assigning the corresponding index as the LBP value (Figure 3.5). This prevents two binary patterns that differ only by a circular shift to be assigned with different indices. For example, the binary patterns "*00010101*", "*10001010*", *01000101*", *10100010*", *01010001*", *10101000*", *01010100*" and *00101010*" will all be encoded with the same scalar. Figure 3.6 shows examples of Local Binary Patterns and Rotation-invariant Local Binary Patterns of a wrinkly eyelid patch.

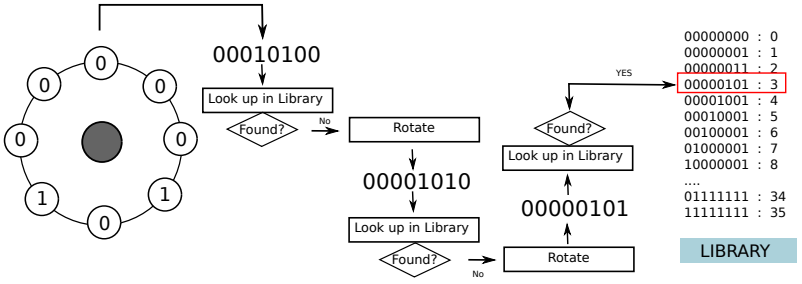


FIGURE 3.5: Rotation-invariant Local Binary Patterns Chart for an 8-pixels Neighbourhood

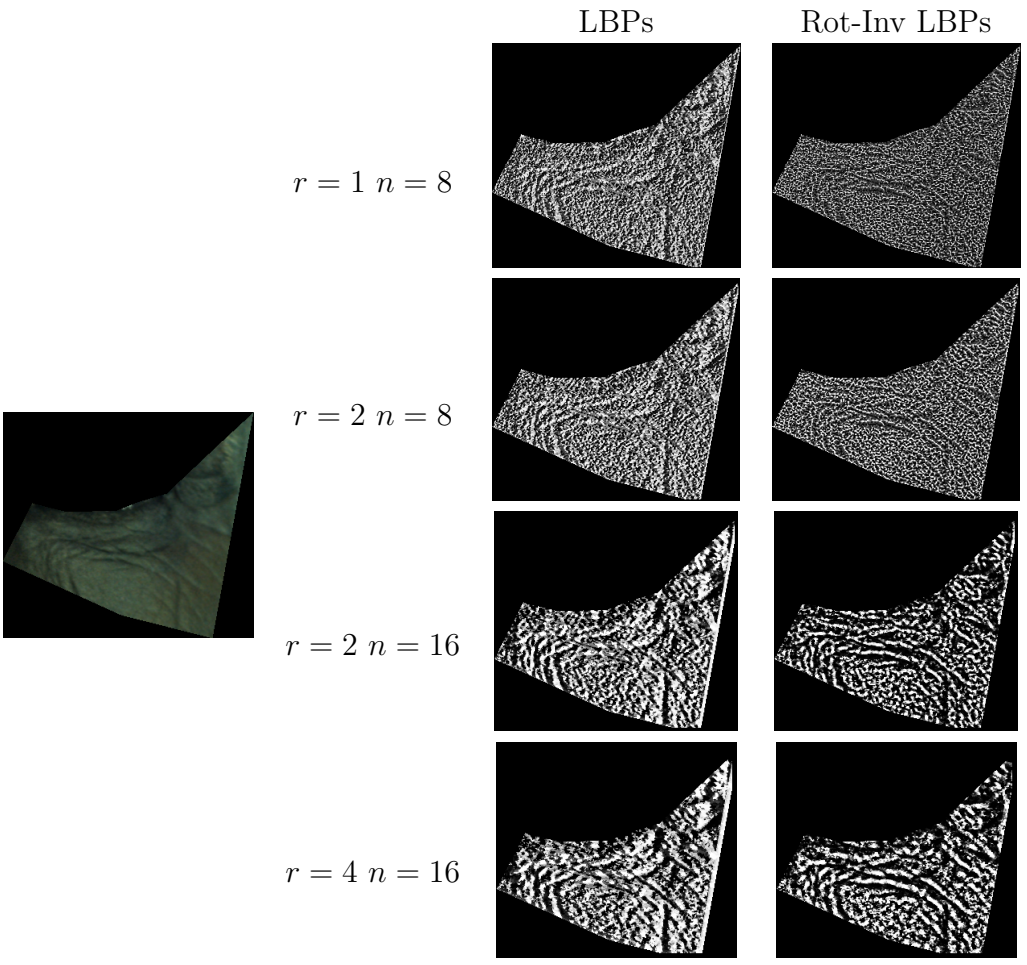


FIGURE 3.6: LBP and rotation-invariant LBPs images of a wrinkly specular albedo patch (r: radius and n: neighbours)

3.4 Simplistic Extensions of the 2D Texture Descriptors to 3D Surface Orientation Data

We extend the three 2D texture measures presented above to analyse 3D orientation texture from dense normal maps. As a normal map is a classical RGB image where the red, green and blue channels encode the X, Y and Z components of the normals respectively, one may consider applying 2D texture analysis algorithms directly to the normal maps. This would be theoretically inconsistent as unlike image pixels, normals do not satisfy the linearity condition required in

the convolutions involved (a linear combination of two normals does not result in a unit-vector). Instead of calculating the texture measures introduced above directly on the normal maps, I propose deriving these from either the slant-tilt space or the tangent space.

3.4.1 Slant-Tilt space

The normal's slant and tilt are extracted at each position (Figure 3.7). This results in a map where at each position we have two values corresponding to the normal's elevation and azimuth. For normalisation's sake, we keep the tangent values so the slant-tilt map is normalized in $[-1, 1]$. Considering $n = (n_x, n_y, n_z)$ denoting a normal, the slant and tilt tangent values are obtained with:

$$\tan\sigma = \frac{\sqrt{x^2 + y^2}}{z} \quad , \quad \tan\tau = \frac{y}{x}$$

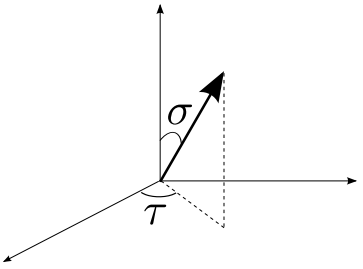


FIGURE 3.7: A normal's slant (σ) and tilt (τ)

3.4.2 Tangent space

In this approach, we consider the normals as elements of a Riemannian manifold and unfold these about the local means using a logarithmic mapping (Figure 3.8).

This results in a tangent map whose elements are 2-dimensional coordinates and are obtained with:

$$\log_{\mu_{\theta_0,\tau_0}}(n_{\theta,\tau}) : \begin{cases} x' = k\cos\theta\sin(\tau - \tau_0) \\ y' = k(\cos\theta_0\sin\tau - \sin\theta_0\cos\theta\cos(\tau - \tau_0)) \end{cases}$$

with $\theta = \frac{\pi}{2} - \sigma$. θ_0 and τ_0 are the spherical coordinates of the local normal mean μ.

At each neighbourhood, the local normal mean is the one that minimises the mean of the geodesic distances to all the other normals in the same neighbourhood.

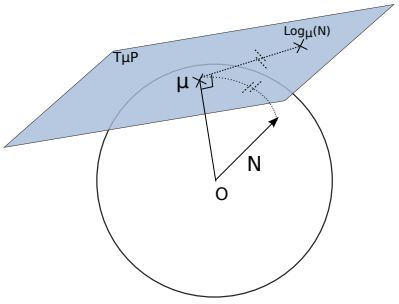


FIGURE 3.8: Projection of a Normal onto the Local Tangent Plane

For visualisation purpose only, we compute a single channel slant/tilt map and tangent map. The pixel values are obtained with (for both the slant/tilt map and the tangent map):

$$I(i,j) = \arctan_2(u,v)\sqrt{u^2 + v^2}$$

u and v represent respectively σ and τ in the slant-tilt case and x' and y' in the azimuthal projection case. Figures 3.9 and 3.10 show respectively the Gabor

responses and LBPs on the tangent and slant/tilt space. One can note that the responses strongly depends on the filter orientation.

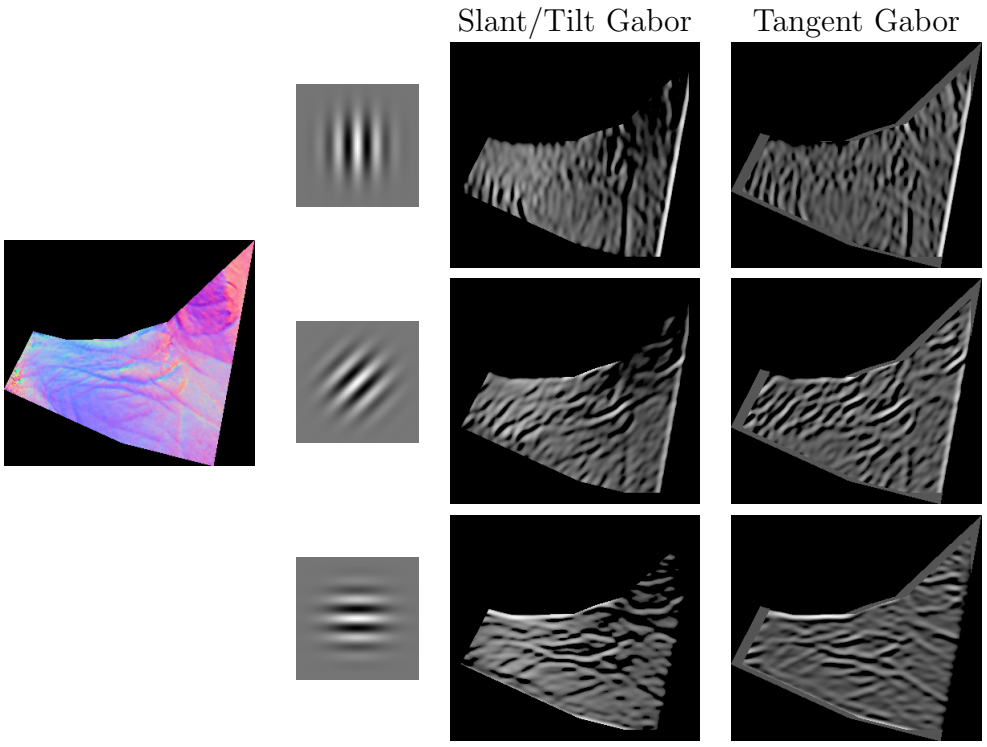


FIGURE 3.9: Gabor responses on slant/tilt and tangent space for different orientations(respectively 0, $\frac{\pi}{3}$ and $\frac{\pi}{2}$) of a wrinkly skin patch surface. The Gabor filter is set with a standard deviation of $6pixels$ and a wavelength of $7pixels$. The images are 310×280

3.4.3 3D Surface Texture Characterization

We adopt a multi-scale scheme where at each level, the texture filter (one of those described in Section 3.3.1) is applied on either the slant/tilt map or the tangent map. As described earlier in this section, both the slant/tilt and tangent map are results of transforming a normal map onto a lesser dimension space. They are both 2D maps with two channels at each pixel. Thus applying the Gabor filter or the LBPs on these results in two responses, one for each channel. The normal map is then down-sampled and passed on to the next level where a new

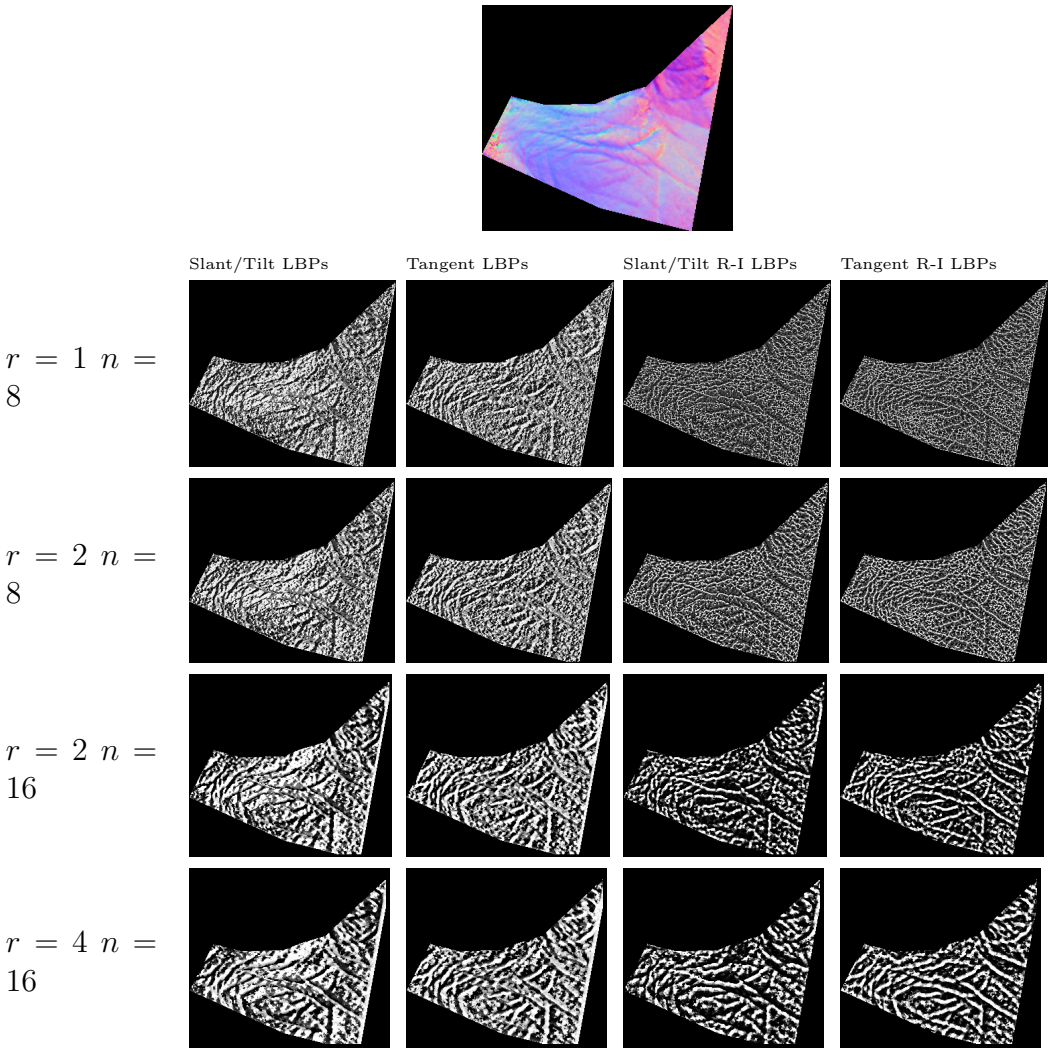


FIGURE 3.10: LBP and rotation-invariant LBPs images of a wrinkly normal map patch in the tangent and slant/tilt spaces (r: radius and n: neighbours)

slant/tilt and tangent map is calculated and the 2D descriptors applied. The process is then repeated for each level of the pyramid.

We normalise all the responses into the interval $[0, 1]$. Assuming $R^{c,l}$ denotes the response on the channel c at the level l, the normalization is performed with:

$$R^{c,l}_{normalized} = \frac{R^{c,l} - \min R^{c,l}}{\max R^{c,l} - \min R^{c,l}}$$

The histograms of the two normalised responses are computed and concatenated

to form the texture descriptor at level l. The same process is repeated at the subsequent level with a down-sampled version of the current normal map.

As illustrated in Section 4.4.3, a convolution should not be done directly on the normals (because of their non-linearity). So the down-sampling is done in the tangent plane with a Gaussian low pass, followed by projecting the result back into the original 3D space using the manifold exponential chart (Section 4.4.3).

3.5 Comparing Classification Results Using the 2D Descriptors and their 3D Extensions

We have run a number of experiments to compare the responses of applying the different texture measures in 2D and 3D configurations and assess their effects on classifying facial skin conditions from high resolution albedo images and the corresponding normal maps.

3.5.1 Feature Extraction

For each sample, we build a 3-level multi-scale feature pyramid. The texture measures described in Section 3.3.1 are used on the albedo samples and their extensions to 3D described in Section 3.4 are used on the corresponding normal map samples. The feature pyramid size depends on the texture measure used and their parameter settings. SVM Ranking is used to reduce the number of features

for all the descriptor to 64. Chapter 5 and 6 give a more detailed presentation of the experimental procedures and dataset.

3.5.2 Discussion

We summarise in Table 3.1 the results of classifying facial wrinkle, large pore and acne presence using 2D texture measures on skin patches albedo and their extensions to 3D surface orientation on the corresponding normal maps.

All three texture characterisation methods show a clear improvement when used in a 3D configuration (slant/tilt or tangent space) on classifying both wrinkle, large pore and acne presence. The improvement is though less important in pore classification, which could be explained by the fact that pores are generally finer than wrinkles, so it is harder to get an accurate capture of their surface topology in a normal map. The Gabor descriptors tend to give better results on classifying wrinkles whereas on classifying pores the Rotation-Invariant Local Binary Patterns appear to achieve better results.

TABLE 3.1: Classification results

		Wrinkles			Pores			Acne		
Features \ Sample Size		20	50	80	20	50	80	20	50	80
2D R-LBPs	$radius = 2$	.53	.59	.62	.61	.63	.62	.59	.63	.60
	$radius = 5$	.60	.67	.70	.73	.73	.72	.62	.70	.70
2D Gabor	$radius = 2$	.60	.65	.72	.58	.58	.59	.60	.62	.61
	$radius = 5$	.64	.70	.75	.71	.70	.71	.71	.73	.71
3D R-LBPs	Slant/Tilt	.75	.78	.81	.79	.81	.80	.71	.73	.72
	Tangent	.70	.73	.79	.73	.75	.73	.66	.69	.65
3D Gabor	Slant/Tilt	**.78**	**.78**	**.82**	**.83**	**.85**	**.85**	**.75**	**.76**	**.77**
	Tangent	.74	.78	.81	.77	.79	.79	.70	.74	.72

3.6 Conclusion

In this section, we have investigated how an extra dimension could add value over classical 2D texture analysis methods, through comparing skin condition classification using 2D texture and 3D surface texture. We have proposed extensions of common 2D texture measures to 3D surface orientation data and conducted a comparative study between these. The results (Chapter 6) show considerable improvements on classifying the presence of facial wrinkles, large pores and acne when extracting the texture feature directly from the normal maps (3D) rather than from the corresponding albedo images (2D).

This, combined to the recent advances in 3D surface recovery techniques and availability of new high resolution 3D dataset, constitutes a motivation to further investigate methods of characterising surface texture in the 3D space. The approaches proposed in this section are simple extensions of two common 2D texture descriptors (LBPs and Gabor filters). In the next chapter, we introduce three novel methods of characterising 3D surface texture directly from a surface orientation field (normal map) and apply these for extracting facial skin microstructure in the three-dimensional space for analysis and/or synthesis purposes.

Chapter 4

Proposed 3D Surface Texture Analysis Methods

4.1 Introduction

As discussed in the previous chapters, the 3D aspect of surface texture has been rather neglected in automated texture analysis. Even though since the early 2000s some studies have addressed the issue [12, 48], a number of works still carry on characterising surface texture of natural objects without taking into account the effects that the interaction between the underlying surface geometry and the imaging conditions (illumination/viewpoint) might have on the apparent texture. The BTF-based methods first introduced in [12] and 3D-texton-based methods introduced in [48], although dealing with the issue by analysing various samples of the same image in different illumination/viewpoint conditions, are

image-based as they do not know the underlying geometry. Also they are known to be very demanding in storage and generally generate extremely high-dimension feature spaces [22].

The previous chapter showed that very simplistic extensions of 2D texture descriptors to 3D data significantly improve classification performances of skin conditions. In this section, we propose and compare three novel geometrical 3D surface texture descriptors that, in contrast to the image-based methods, operate directly on the surface geometry and produce more compact representations.

4.2 Proposed Method I: Rotation Fields Pyramid

4.2.1 Introduction

In this section we propose an approach based on multi-resolution rotation fields. Rotation fields are a means of capturing high frequency information from surface orientation. [60] employed these to correct the 3D position of 3D mesh vertices with accurate high frequency data from normal map measured with photometric stereo. Frequency separation has been extensively used in the literature to represent 2D texture [2, 36, 69]. This generally involves a pyramidal multi-resolution representation which allows capturing texture information at different scales. At each level of the pyramid, the low frequency information is separated from the

high frequency; while the former is more related to global shape, the latter can be a representation of the local texture.

We propose a multi-resolution analysis scheme where at each level of the pyramid, the low frequency information in the normal map is separated from the high frequency in the form of rotation fields. We also demonstrate that the rotation fields pyramid can be used not only on analysis but also to synthesise 3D surface texture.

4.2.2 Rotation Fields

Let N denote a normal map and $N_{i,j}$, the normal vector at the pixel $p_{i,j}$, we compute a smoothed version N^s of N by computing at each pixel a weighted geodesic or euclidean mean over a neighbourhood Ω with a radius r (a post-normalisation of the resulting normal is required in case of the euclidean mean). The weights $w_{i,j}$ are determined by a Gaussian with a standard deviation the same as the radius of the neighbourhood. The geodesic mean is defined as:

$$\mu = argmin_{N'} \sum_{p_{i,j} \in \Omega} d(N_{i,j}, N') \tag{4.1}$$

with $d(N_{i,j}, N')$ being the geodesic distance between $N_{i,j}$ and N'). [65] show that this can be recursively approximated by:

$$\mu^{t+1} = exp_{\mu^t} \left(\frac{1}{card(\Omega)} \sum_{\Omega} log_{\mu^t} \left(N\left(i,j\right) \right) \right) \tag{4.2}$$

If introducing the Gaussian weights $w_{i,j}$, obtain:

$$\mu^{t+1} = exp_{\mu^t}\left(\frac{1}{card(\Omega)}\sum_{\Omega} w_{i,j} log_{\mu^t}\left(N\left(i,j\right)\right)\right) \tag{4.3}$$

(exp_{μ^t} and log_{μ^t} are the exponential and logarithm map about the geodesic mean μ_t). The rotation fields R are obtained by computing, at each pixel, the rotation to apply on the original normals to match the smoothed ones. An axis-angle representation $[\vec{e}, \theta]$ can be adopted to characterise each rotation with four parameters (three for the axis $\vec{e}$ and one for the angle θ). Denoting $R^{\vec{e}}$ as the axis component of the rotation field and R^{θ} as the angle component, gives:

$$R^{\vec{e}}_{i,j} = N_{i,j} \times N'_{i,j} \quad \text{and} \quad R^{\theta}_{i,j} = N_{i,j} \cdot N'_{i,j} \tag{4.4}$$

The rotation axis, $R^{\vec{e}}_{i,j}$, can be normalized (dividing by it's norm) so that the rotation parameters can be brought down to three:

$$R_{i,j} = \frac{R^{\vec{e}}_{i,j}}{\|R^{\vec{e}}_{i,j}\|} R^{\theta}_{i,j} \tag{4.5}$$

For visualisation purposes we encode these three parameters in an RGB image. The smoothing radius r controls the level of detail extracted. Small values of r allow extracting very fine skin texture (down to the level of the pores) while higher values tend to capture medium frequency structures such as acne and

wrinkles. Figure 4.1 shows the rotation maps and corresponding low frequencies of a wrinkly normal map patch computed with three different radius values.

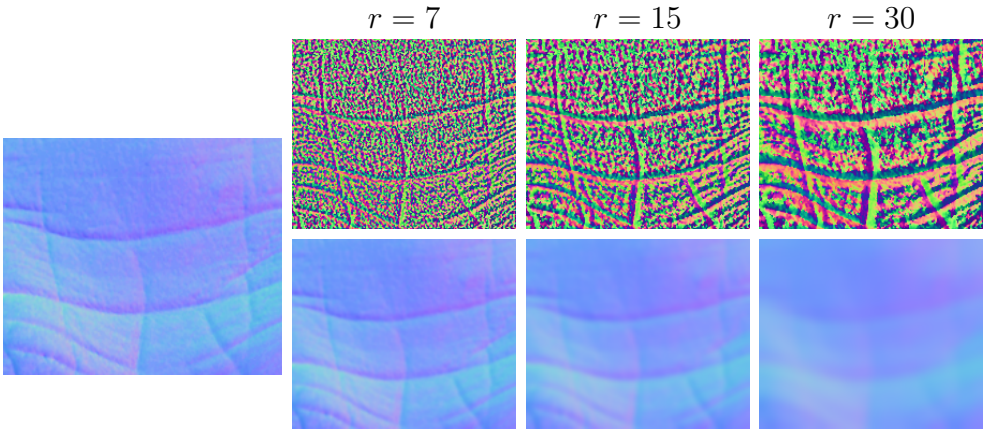


FIGURE 4.1: Rotation maps (top row) and corresponding low frequencies (bottom row) with different radius values ($r = 7, 15, 30$) of a wrinkly normal map patch

4.2.3 Rotation Fields Pyramid

Given a normal map N, a subband decomposition can be performed by building an image pyramid, where at each level the low frequency information is separated from the high frequency. The whole process is started by applying a low pass filter G_0, namely a Gaussian (with geodesic or euclidean averaging). The result is a normal map L_0 representing the low frequency surface variation of the original one. This is then used to extract the high frequency information by calculating the rotation field that brings it back to the original normal map. After extracting the high frequencies in the form of rotation field H_0, the low frequency normal map L_0 is then down-sampled and passed on the next level where the same process is repeated (Figure 4.2). H_n and R_n have the same meaning which the rotation field extracted at the level n of the pyramid. On the synthesis side, the

reconstruction can be initialised by the low frequency component of any normal map, but it is important that the low frequency is produced after at least the same number of decomposition level as the original.

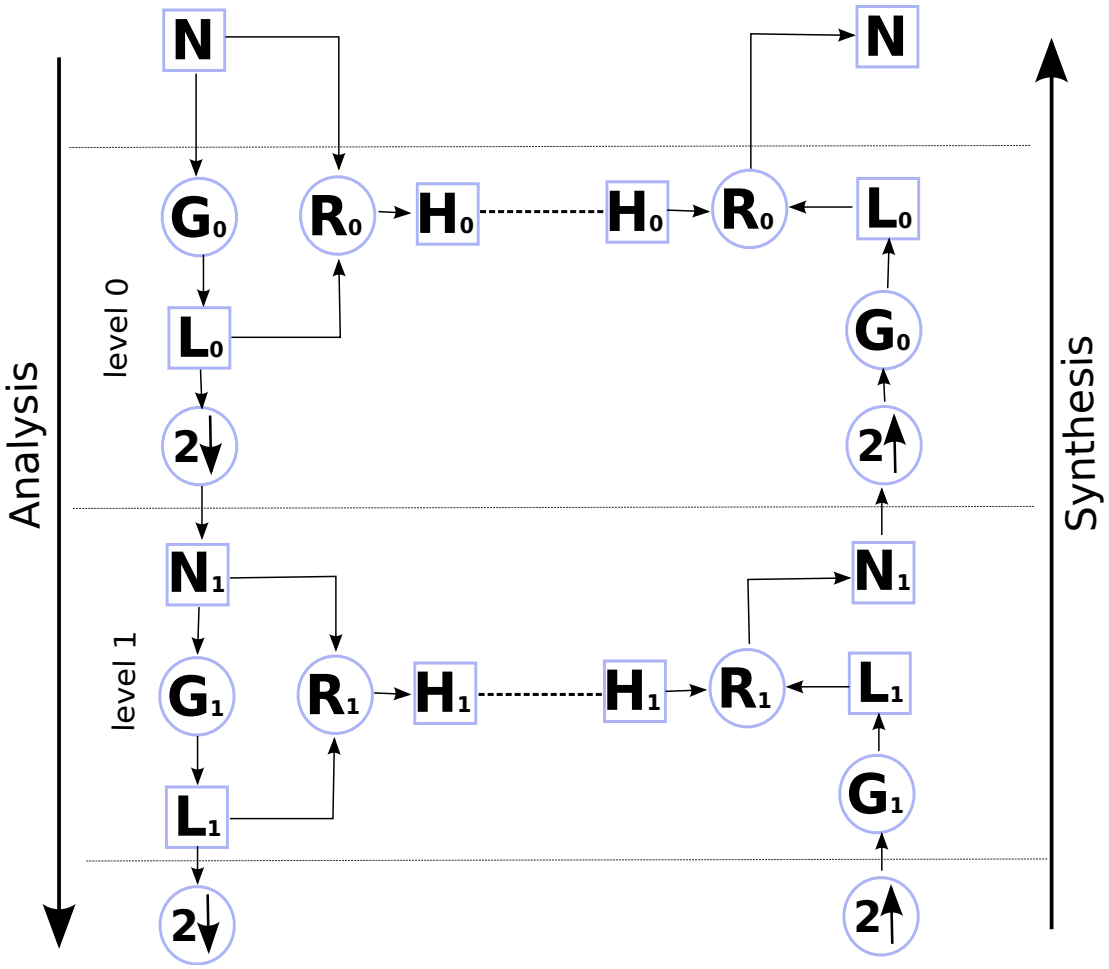


FIGURE 4.2: Rotation fields pyramid

In the two-dimensional case, most of the studies that use a pyramidal representation extract the high frequency information in several subbands. The main motivation of this is to capture different spatial configurations and orientations of the texture. For example [36] employed steerable filters to capture anisotropic texture with the presence of elongated or oriented structures. However, in contrast to individual pixels in 2D images, each surface orientation in the normal map encodes a certain level of information about the surface gradient within its

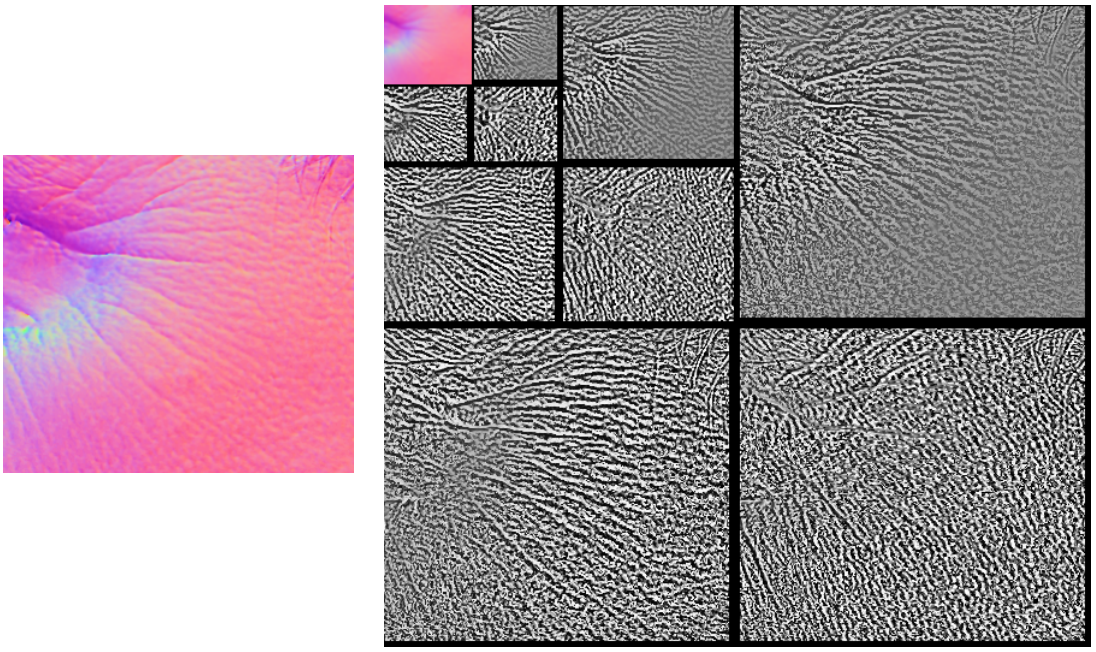


FIGURE 4.3: A 3-level rotation fields pyramid of a wrinkly normal map patch

immediate neighbourhood. So, at each level of the the pyramid, we use three subbands that correspond respectively to the three components of the rotation vector. Figure 4.3 shows a 3-level rotation field pyramid of a wrinkly normal map patch.

4.2.4 Riemannian Distance on the Rotations Group SO_3

After having represented the 3D surface texture as an $n-$level pyramid of rotation fields, a metric in the rotation space is needed in order to analyse their spatial distribution. In other words a way to compare two rotation vectors is needed. This problem has been studied in [64].

Beside the angle-axis representation, rotations can be represented by 3×3 orthogonal matrices which form the Rotation Group SO_3 and constitute a smooth manifold [64]. This means that the set of rotation matrices is differentiable and

supports a Riemmannian metric allowing to compute the distance between rotations. If $\mathcal{R}_1$ and $\mathcal{R}_2$ are two rotation matrices R_1 and R_2 the corresponding axis-angle representations (the conversion can be easily done with the Rodriguez formula), the Riemmannian distance between $\mathcal{R}_1$ and $\mathcal{R}_2$ is given by [64]:

$$d: \begin{array}{ccc} SO_3 \times SO_3 & \rightarrow & \mathbb{R} \\ (\mathcal{R}_1, \mathcal{R}_2) & \mapsto & d(\mathcal{R}_1, \mathcal{R}_2) = (R_2 \circ R_1) \end{array} \tag{4.6}$$

Denoting the composition of rotations R_1 and R_2. Whilst the composition of rotations can be calculated by the dot product of the two matrices ($R_2 \circ R_1 \sim \mathcal{R}_2^T \cdot \mathcal{R}_1$), [64] showed that it is more advantageous to use unit quaternions as an intermediate step because the result is easier to differentiate. The idea is to convert the axis-angle representation of the rotations to unit quaternions, multiply these and convert back into axis-angle representation. Let R be an axis-angle rotation (axis denoted by $R^{\vec{e}}$ and angle by R^{θ}) and its corresponding unit quaternion Q represented by its scalar s and vector v parts, the conversions are given by:

$$Q: \begin{cases} s = \cos\frac{R^{\theta}}{2} \\ v = R^{\vec{e}} \sin\frac{R^{\theta}}{2} \end{cases} \tag{4.7}$$

$$R: \begin{cases} R^{\theta} = 2atan_2(|v|, s) \\ R^{\vec{e}} = \frac{v}{\sin\frac{\theta}{2}} \end{cases} \tag{4.8}$$

And for two unit quaternions $Q_1(s_1, v_1)$ and $Q_2(s_2, v_2)$ the non-commutative multiplication is given by:

$$Q_1 * Q_2 : \begin{cases} s_1 s_2 - v_1 \cdot v_2 \\ s_1 v_2 + s_2 v_1 + v_1 \times v_2 \end{cases} \tag{4.9}$$

Equations 4.6, 4.8 and 4.9 give:

$$d(\mathcal{R}_1, \mathcal{R}_2) = atan_2 \begin{pmatrix} |s_1 v_2 + s_2 v_1 + v_1 \times v_2| \\ s_1 s_2 - v_1 v_2 \end{pmatrix} \tag{4.10}$$

Replacing 4.7 in 4.10 yields:

$$d(\mathcal{R}_1, \mathcal{R}_2) = atan_2 \begin{pmatrix} \left| R_2^{\vec{e}} \cos\frac{R_1^\theta}{2} \sin\frac{R_2^\theta}{2} + R_1^{\vec{e}} \cos\frac{R_2^\theta}{2} \sin\frac{R_1^\theta}{2} + \left(R_1^{\vec{e}} \times R_2^{\vec{e}}\right) \sin\frac{R_1^\theta}{2} \sin\frac{R_2^\theta}{2} \right| \\ \cos\frac{R_1^\theta}{2} \cos\frac{R_2^\theta}{2} - \left(R_1^{\vec{e}} \cdot R_2^{\vec{e}}\right) \sin\frac{R_1^\theta}{2} \sin\frac{R_2^\theta}{2} \end{pmatrix} \tag{4.11}$$

For the sake of simplicity, we assume that the two rotation axes $R_1^{\vec{e}}$ and $R_2^{\vec{e}}$ are parallel. This simplifies Equation 4.11 to:

$$d(\mathcal{R}_1, \mathcal{R}_2) = atan2 \begin{pmatrix} \sin\left(\frac{R_1^\theta}{2} + \frac{R_2^\theta}{2}\right) \\ \cos\left(\frac{R_1^\theta}{2} + \frac{R_2^\theta}{2}\right) \end{pmatrix} \tag{4.12}$$

In our application, this simplification do not alter the captured information in terms of surface irregularities. Indeed on the rotation map, each pixel represents the rotation of the original surface normal from the smooth one, hence characterised by two parameters: the axis and the angle of rotation. The angle quantifies how much the two normals deviate from each other whereas the axis

determines the plane in which the rotation happens. Now when we compute the distance between two rotations, we are more interested in capturing the deviation component than the orientation component of the rotation. This led us to assume that the two rotations have the same axis which considerably simplifies the calculation without losing the deviation information we want to capture.

4.2.5 Feature Extraction

For a given normal map patch, we compute an $l-$level rotation field pyramid. For each level of the pyramid, we compute at each pixel the distances between the corresponding rotation and each of the neighbouring pixels within an $N \times M$ neighbourhood using Equation 4.12. This gives at each pixel a vector of length $N \times M - 1$. We complete this vector with the rotation vector at the central pixel yielding a vector $N \times M + 2$ long. As we want to end up with a histogram as feature vector, a vector quantisation algorithm is used to map each of these $N \times M + 2$ vectors to a scalar value. In this work, we use K-means [51] which introduces another parameter k representing the number of clusters. Each cluster is associated to a symbolic label (a scalar value). We then compute the histogram of the resulting map of symbolic labels. The size of the histogram is given by the number of clusters k. The process is repeated at each level in the pyramid and the histograms from all the levels are concatenated to form the $l \times k$ feature vector associated with the patch.

4.2.6 Classification Results

The method is tested on classifying the three skin conditions we are interested in from our collected 3D facial dataset. We experimented with two different k-mean configurations, $k = 100$ and $k = 200$, yielding respectively, with a 3-level pyramid, 300 and 600 long feature vectors. SVM Ranking is then used to reduce both feature vectors to 64 components. Also various sample sizes were tried: 20×20 pixels, 50×50 pixels and 80×80 pixels. Chapters 5 and 6 give a more detailed presentation of the dataset and experimental setup. The classification results given in Table 4.1 show that higher values of k, the number of clusters, yield better classification performances. This can be explained by the fact that higher values of k means higher resolutions in the quantisation which enables losing less information after the K-means is performed. In Chapter 6, these results are compared against the other proposed descriptors and against a BTF based method.

TABLE 4.1: Classification results

		Wrinkles			Pores			Acne		
Features \ Sample Size		20	50	80	20	50	80	20	50	80
Rot. Fields	$K = 100$	.78	.80	.84	.87	.89	.87	.79	.83	.83
	$K = 200$	.82	.86	.90	.91	.90	.92	.84	.87	.87

4.2.7 Potential Application: 3D Skin Texture Synthesis and Transfer

Beside its ability to produce discriminative descriptors for three-dimensional surface texture classification, the rotation fields pyramid can also be used to synthesise or transfer three-dimensional surface texture. Many studies have used image pyramidal representation to synthesise two-dimensional textures [2, 36, 69]. [36] used histogram matching through all the levels of a steerable pyramids [27] computed from the source and target texture [36]. [69] proposed a parametric method which constrains the target pyramid to satisfy a set of joint statistics computed on the pyramid basis functions.

Realistically simulating 3D skin texture has received lot of interest in the Computer Graphics community. One of the earliest works is [28] which built a statistical model of facial 3D meso-structures such as wrinkles. They have captured high resolution facial scans, separated the skin details from smooth meshes in terms of depth variation and used the pyramid-based texture synthesis proposed in [36]. However this methods, beside capturing only the skin meso-structures, suffers from lack of regularity in the synthesised texture due to the statistical nature of the synthesis method [36]. More recent and advanced works on 3D skin texture synthesis include [30, 59]. These also model the 3D skin roughness in terms of displacement maps. However they use a more advanced capture device to acquire 3D skin details at the micro-scales which allowed them to synthesise the skin micro-structure.

We demonstrate here that the Rotation Fields can be a potential alternative to the displacement maps which are the main representation used for synthesis to date. Various 2D texture synthesis techniques can be extended to 3D surface orientation data using the rotation fields. Given an example normal map, the idea is to synthesise a new normal map that presents similar surface appearance. In the 2D case, the texture is assumed to be unaffected by shape and is generally synthesised directly from the measured albedo. However in the 3D case, dense normal fields encode the surface fine texture and shape at the same time. This shape-related information present in the normal map needs to be factorised out in order to synthesis only the texture-related information. This implies an additional transfer step which consists in combining the synthesised 3D surface texture and the target's overall shape.

We propose a unified approach where the inputs are two normal maps: a *source* from which we extract the texture to synthesise and a *target* to which we transfer the synthesised texture. Figure 4.4 summarises the synthesis and transfer processes. We first separate the high frequency information (texture) from the low frequency (shape) in both the *source* and *target* normal maps (the separation process is explained in Section 4.2.2). The high frequency from the *source*, kept in the form of rotation fields, is then grown to match the *target* size using one of the methods in Section 4.2.8. The final result is produced by deforming the low frequency from the *target* using the synthesised and grown rotation field.

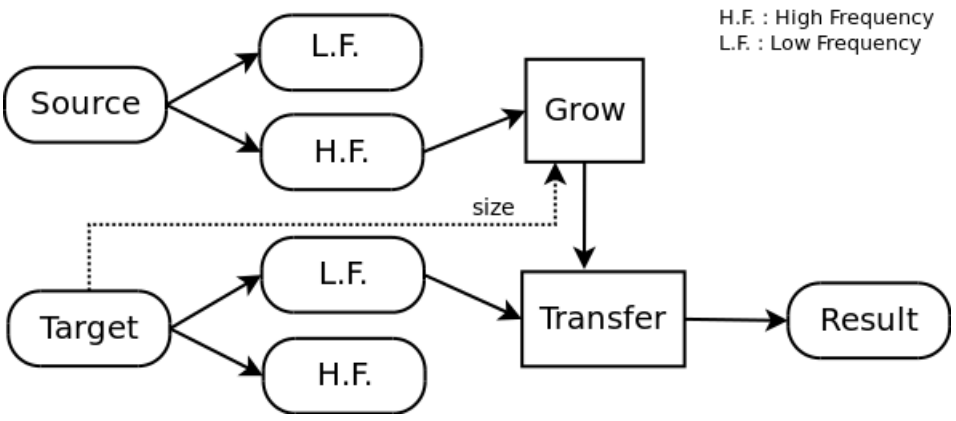


FIGURE 4.4: Flowchart of the synthesis and transfer process

4.2.8 Growing the Rotation Field

Various methods of growing two-dimensional texture have been proposed in the literature. The most popular ones include pixel-based synthesis and patch-based synthesis.

In the pixel-based methods, one pixel is generated at a time during the synthesis process. One of the reference works is due to [20], who propose to synthesise each pixel in the target texture by comparing its current neighbourhood with all possible neighbourhoods in the source image. To introduce some randomness, instead of choosing the closest target neighbourhood, the authors randomly pick a neighbourhood from the $n-$closest. [87] propose avoiding the exhaustive search in Efro and Leung's method by quantising the neighbourhood search space using a tree-structured vector quantisation.

In the patch-based methods, texture is synthesised by copying patches from the source and seamlessly stitching these together into the target. The patches are copied from the top-left to the bottom-right of the target. At each copy, a patch is chosen from the source according to a similarity measure between

its top/left borders and the borders of the already copied patches adjacent to the current position. Different approaches for stitching the copied patches have been proposed in the literature. [19] used dynamic programming to compute a minimum error boundary cut between overlapping patches. [47] propose a graph cut technique which has the advantage of automatically finding the optimal patch size to copy from the source, in contrast to Efros and Freeman's method, where the patch size is fixed.

In the next three sections, we extend these different two-dimensional texture growing techniques to the proposed rotation field representation in order to synthesise 3D surface texture.

4.2.8.1 Pixel-based Texture Synthesis

We extend the algorithm in [20] to the proposed rotation fields by using the metric defined in Equation 4.12 to compute the distance between two rotations. Given a normal map patch, the synthesis process starts by extracting the high frequency information into a rotation field. The extracted rotation field is then grown to match the target patch's size using our extended version of Efros and Leung. At the same time the target normal map is down-sampled with the same window size used to extract the high frequency from the source normal map. The result, which represents the low frequency information in the target, is then corrected using the synthesised rotation field.

4.2.8.2 Patch-based Texture Synthesis

Here, we use an image quilting method, introduced in [19], to grow the rotation field. The approach consists in copying patches from the source rotation field to the target in such a way that they overlap. Then dynamic programming is used to find, in each overlapping region, the best cut that gives seamless transition between the patches involved. In the original formulation, the minimum error cut between 2D image-patches is computed by solving the following dynamic programming problem:

$$E_{i,j} = e_{i,j} + \min\left(E_{i-1,j-1}, E_{i-1,j}, E_{j+1}\right) \tag{4.13}$$

where $e_{i,j}$ represents the error surface at the pixel (i, j) in the overlapping region. It is computed as the square difference of corresponding pixel values from the overlapping patches. In this case, we compute $e_{i,j}$ using the rotation distance defined in Equation 4.12.

$$e_{i,j} = d(\mathcal{R}^1_{i,j}, \mathcal{R}^2_{i,j}) \tag{4.14}$$

4.2.8.3 Warping-based Texture Synthesis

The pixel-based and patch-based growing methods can both be used to synthesise arbitrary sized target textures. However they both involve expensive search and therefore have limited potential for tasks such as real-time synthesis. The warping

based methods grow the source texture to match a target according to a certain number of pre-defined control points [82]. Concretely, these approaches consist in placing a set of control points in the source texture and the target space. Each control point in the source is associated to one in the target space. The growing process finds the transformation that maps the source's control points to the target's in the texture coordinate space and applies image warping techniques to interpolate between the source texture and target.

While the warping based methods are not adequate when the size ratio between target and source is considerable, they have the advantage of being fast and conserving the visual regularity of the texture.

In this study, we aim at synthesising by example skin conditions such as wrinkles, acne or pores which generally have a certain visual regularity in how they are distributed across the face. So it is important that the growing method we use conserves this regularity. Between the three growing methods described above, the pixel based methods are the least regular as shown in Figure 4.5. These are appropriate for synthesizing rather stochastic textures. The patch based methods tend to conserve more regularity that the pixel based methods, but this still depends on the chosen patch size. The bigger the size, the more regular the output will be.

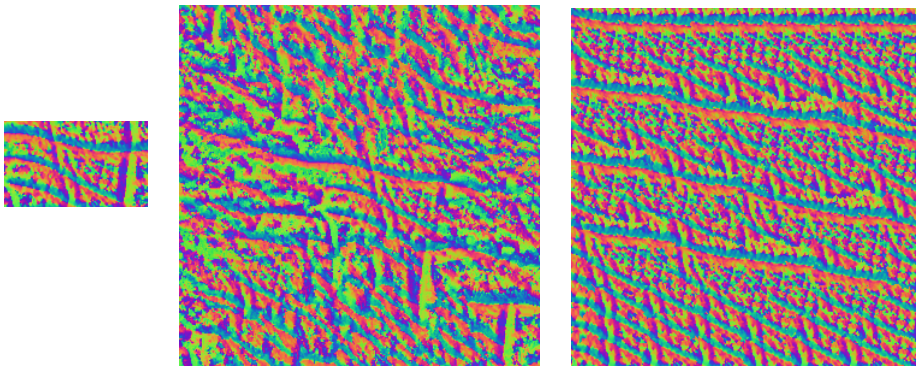


FIGURE 4.5: Rotation field growing

4.2.9 Multi-scale Synthesis with the Rotation Field Pyramids

We use the Thin Plate Spline [17] with overlap warping method proposed in [82] to transform the source rotation field to the target configuration. The transformation is governed by a set of control points placed in the source and target. These are associated to each other using visual cues (e.g. a control point on an eye-corner in the source is associated with a control point on the corresponding eye-corner in the target). In fact these control points are picked from the template used to segment the face in different regions (in Section 5.3.2.4).

Concretely, to transfer the forehead skin texture from the subject A to the subject B in Figure 4.6, we use the forehead masks in the 2D texture space for each of the three poses (how these masks are obtained is explained in Section 5.3.2.4) to select the regions of interest in the normal map of each pose. For the source (subject A), the pyramidal rotation field is computed in the region of interest (the rotation field pyramid calculation is presented in Section 4.2.3) and for the target (subject B), the low frequency surface normals are computed by Euclidean

Gaussian or Weighted Geodesic mean. Then the source rotation field pyramid is grown to match the target using a thin plate spline with overlap warping method to match the target's size and mask. The warping is controlled by the projected 2D landmarks. Finally, the target's low frequency normals are corrected at each level of the pyramid by rotating them using the warped rotation field.

4.2.10 Transfer Across the Whole Face

Any region of the face can be considered individually to transfer skin texture or a combination of regions can be used to transfer the texture across the whole face. Moreover, choosing different texture sources for different regions allows synthesising facial appearance combining skin conditions from different individuals.

As the synthesis is done on the normal maps, in the texture space, to visualise the result the mesh is rendered by shading the specular component using the synthesised normals. However the synthesised normals can be used to refine the position of the mesh's vertices so that the synthesised 3D texture is embedded directly in the geometry (Figure 4.7). This can be realised using the position refinement algorithm proposed in [60] where the mesh vertices position are refined by optimizing the error between the measured normals and the mesh normals.

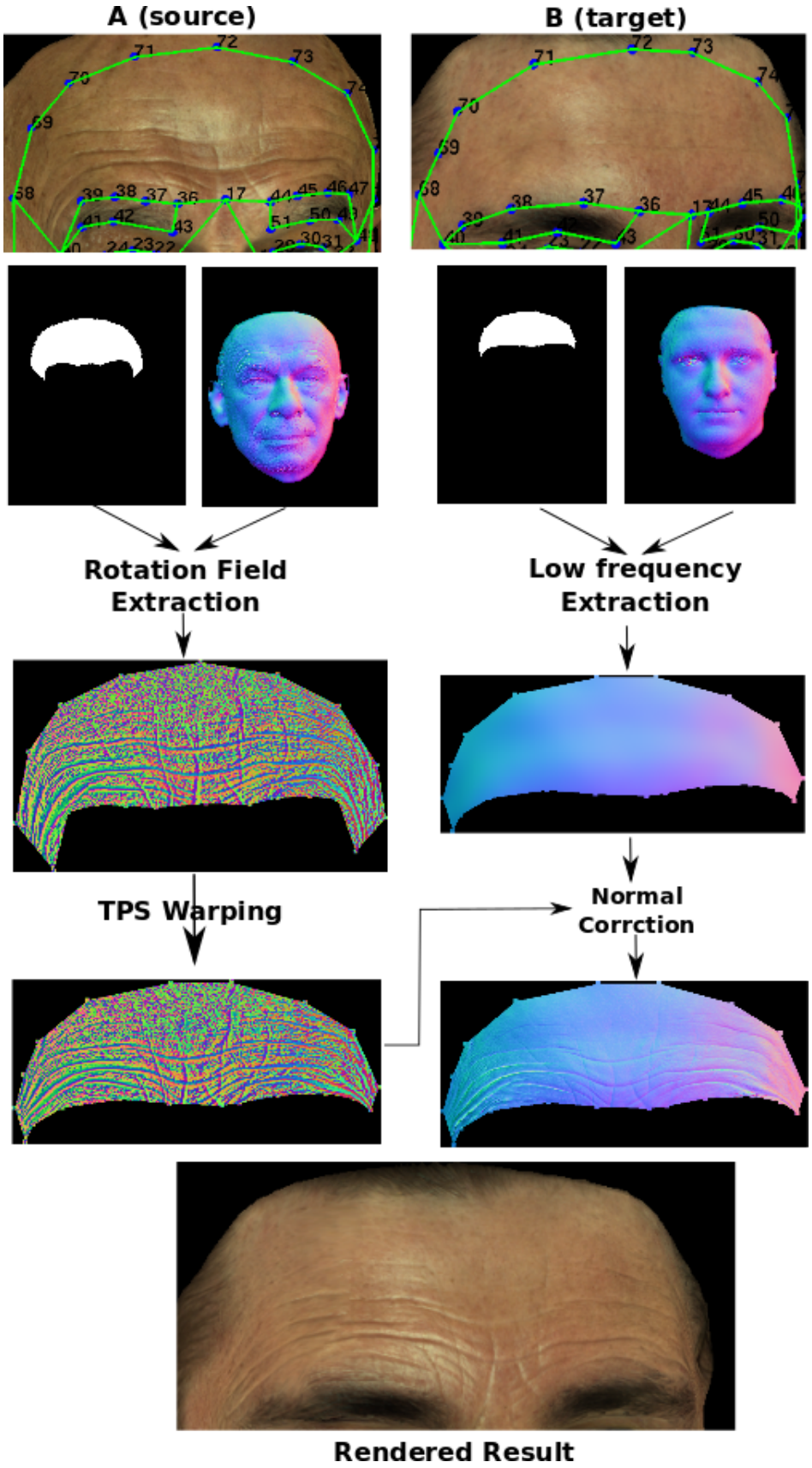


FIGURE 4.6: Synthesis and transfer of 3D skin texture on forehead

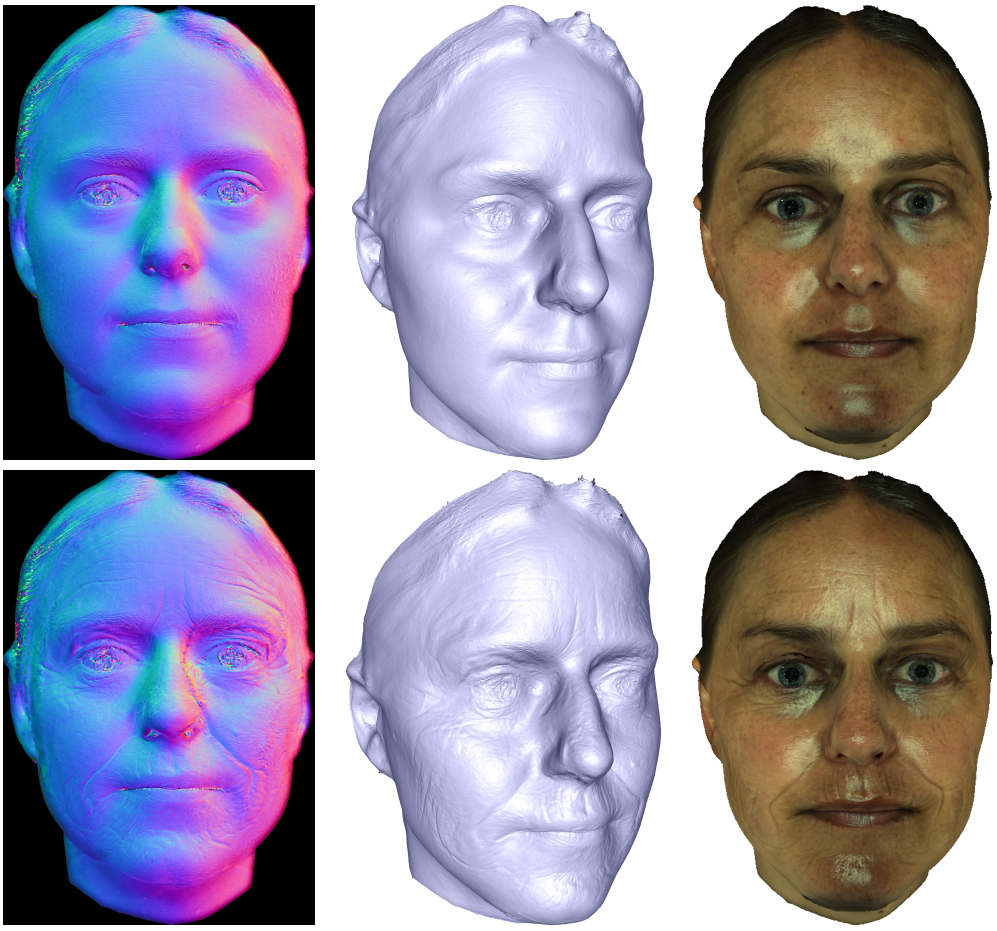


FIGURE 4.7: Synthesis across the whole face and mesh refinement

4.2.11 Conclusion

In this section a multi-resolution rotation field for characterising 3D surface texture was presented. The method consists in separating the high frequency information present in a normal map from the low frequency information at different scales. The high frequencies are encoded in the form of a rotation field pyramid, whereas the residual low frequency is a smoothed normal map. The rotation field pyramid corresponds to a set of rotations to apply to the residual low frequency normal map at each pixel to recover the original normals. The representation can be used to analyse as well as to synthesise 3D skin texture. The analysis consists in extracting a histogram of features from the rotation field pyramid to

characterise the texture (Section 4.2.5) and feeding it into a machine learning algorithm (we give the experimental results in Chapter 6). The synthesis, as presented in Section 4.2.7, consists in growing the rotation map of a skin texture example and combining the result with the low frequencies of the target.

4.3 Proposed Method II : Local Orientation Patterns

4.3.1 Introduction

In this section, we propose an approach based on the notion of texture spectrum which characterises 2D textures by encoding the spatial variations of pixels in a neighbourhood and was introduced in the early nineties [85]. We extend this to 3D surface data and propose two methods of assessing the variation of orientation between two normals. The first applies a threshold to the normals' dot product, while the second variant directly compares their polar and elevation angles resulting in a four-level texture unit.

4.3.2 Background: Texture Spectrum

The notion of *texture spectrum* was introduced in [85] and can be defined as the distribution of texture entities called *texture unit* over an image. In the original formulation, a *texture unit* is a 3×3 pixels neighbourhood forming a window of 8

pixels $(p_i)_{1\leq i\leq 8}$ surrounding a central one p_0. Each of the eight surrounding pixels may be associated to three possible patterns defined by the function $(f_i)_{1\leq i\leq 8}$:

$$(f_i)_{1\leq i\leq 8} = \begin{cases} 0 & \text{if} \quad p_i < p_0 \\ 1 & \text{if} \quad p_i = p_0 \\ 2 & \text{if} \quad p_i > p_0 \end{cases} \tag{4.15}$$

The value of the texture unit associated to p_0 is determined from the 8 surrounding the patterns by:

$$f(p_0) = \sum_{i=1}^{8} f_i \times 3^{i-1} \tag{4.16}$$

A texture unit is associated to each pixel contained in the image and the texture spectrum is defined as the distribution of texture units over the whole image. This is represented by a histogram keeping the frequency of each possible texture unit value over the image.

Texture spectrum based approaches have proven their powerful ability to discriminate texture variations in a 2D image, despite their simplicity and ease of use. Local Binary Patterns [63, 85] are an extension of the texture spectrum model where the number of patterns is brought down to two and the neighbourhood radius, shape and pixels number can be parameterised.

4.3.3 Generalising Texture Spectrum

The notion of texture spectrum can be generalised by extending the definition of a texture unit to n possible patterns between two pixels and an arbitrary number of pixels, N, uniformly surrounding a central pixel, p_0, with an arbitrary radius of r. In this case the texture unit function (Equation 4.16) becomes:

$$f(p_0) = \sum_{i=1}^{N} f_i \times n^{i-1} \tag{4.17}$$

The patterns $(f_i)_{1 \leq i \leq N}$ can be defined with any discrete two dimensional function that has only n possible values in $\mathbb{Z}^+$. Assuming that the pixel values are picked from a set $\mathbb{E}$ and that $\mathbb{A}$ denotes the set of n possible pattern values, $(f_i)_{1 \leq i \leq N}$ is given by:

$$\begin{array}{rccccl} (f_i)_{1 \leq i \leq N} : & \mathbb{E} \times \mathbb{E} & \rightarrow & \mathbb{A} & \text{with:} \quad \mathbb{A} \subset \mathbb{Z}^+ \quad \text{and} \quad |\mathbb{A}| = n \\ & (p_0, p_i) & \mapsto & f_i(p_0, p_i) & \end{array} \tag{4.18}$$

As the final objective is to get a histogram representing the distribution of texture units over the image, it is important to choose the pattern function $(f_i)_{1 \leq i \leq N}$ well as its number n of possible values and the chosen number of neighbours determines the number of bins to reserve for the histogram:

$$B = n^N$$

Finally, with a fixed radius and number of neighbours, each neighbour position (x_i, y_i) is obtained by:

$$x_i = round(r\cos(\frac{2\pi i}{N})) \quad \text{and} \quad y_i = round(r\sin(\frac{2\pi i}{N}))$$

4.3.4 Local Orientation Patterns

The approach we propose for analysing 3D surface texture from normal maps is entirely based on the generalised Texture Spectrum. The main task here is to find good pattern functions that can represent the normals' orientation distribution over a texture unit. As each pixel of a normal map is an element of $[-1,1]^3$, we are looking at pattern functions of the form:

$$(f_i)_{1\leq i\leq N}: \begin{array}{ccccc} [-1,1]^3\times[-1,1]^3 & \to & \mathbb{A} & \text{with:} & \mathbb{A}\subset\mathbb{Z}^+ \quad \text{and} \quad |\mathbb{A}|=n \\ (\mathcal{N}_0,\mathcal{N}_i) & \mapsto & f_i(\mathcal{N}_0,\mathcal{N}_i) & & \end{array} \tag{4.19}$$

with $\mathbb{A}$ the set of possible pattern values $\mathcal{N}_0$ the central normal and $\mathcal{N}_i$ one of the N surrounding normals.

We propose two pattern functions for representing the normals' orientation distribution. The first function computes the dot product of two normals and compares the result with a threshold. The second function compares the azimuthal and polar angles of the normals directly.

4.3.4.1 First Pattern Function

The first pattern function we propose evaluates the dot product between the central normal and one of the surrounding normals, and compares the result to a threshold. Formally it is given by (with a threshold τ):

$$f_i^{\tau}(\mathcal{N}_0, \mathcal{N}_i) = \begin{cases} 0 & \text{if} \quad \mathcal{N}_0.\mathcal{N}_i < \tau \\ 1 & \text{if} \quad \mathcal{N}_0.\mathcal{N}_i \geq \tau \end{cases} \tag{4.20}$$

With this pattern function, the number of bins needed for the histogram is given by 2^N as in Local Binary Patterns. As the normals are normalized in $[-1, 1]$, the dot product depends only on the angle between the two normals. However the problem here is to find a threshold.

The threshold: It is clear that a good threshold depends on the local orientation distributions in the normal map; a good threshold for a dense and/or more or less uniform normal map may not be suitable for a sparser normal map. The threshold choice also depends on the application; for the same normal map, we may use different thresholds depending on whether we want to capture high or low-frequency variations (although this would need to be combined with an adequate radius setting).

We tried two techniques for choosing the threshold. The first averages the dot products of all pairs of normals. The second method computes a threshold map by locally averaging the dot products between each normal in a Texture Unit

with the central normal. Our experiments show that the first method achieves better results than the second (Chapter 6), although we believe that a good threshold map may provide additional robustness in cases where the distribution of the normal orientations varies considerably from one place to another.

Figure 4.8 shows the Local Orientation Pattern Images of three skin patches using the first patterns function with a radius of 1, 2 and 4.

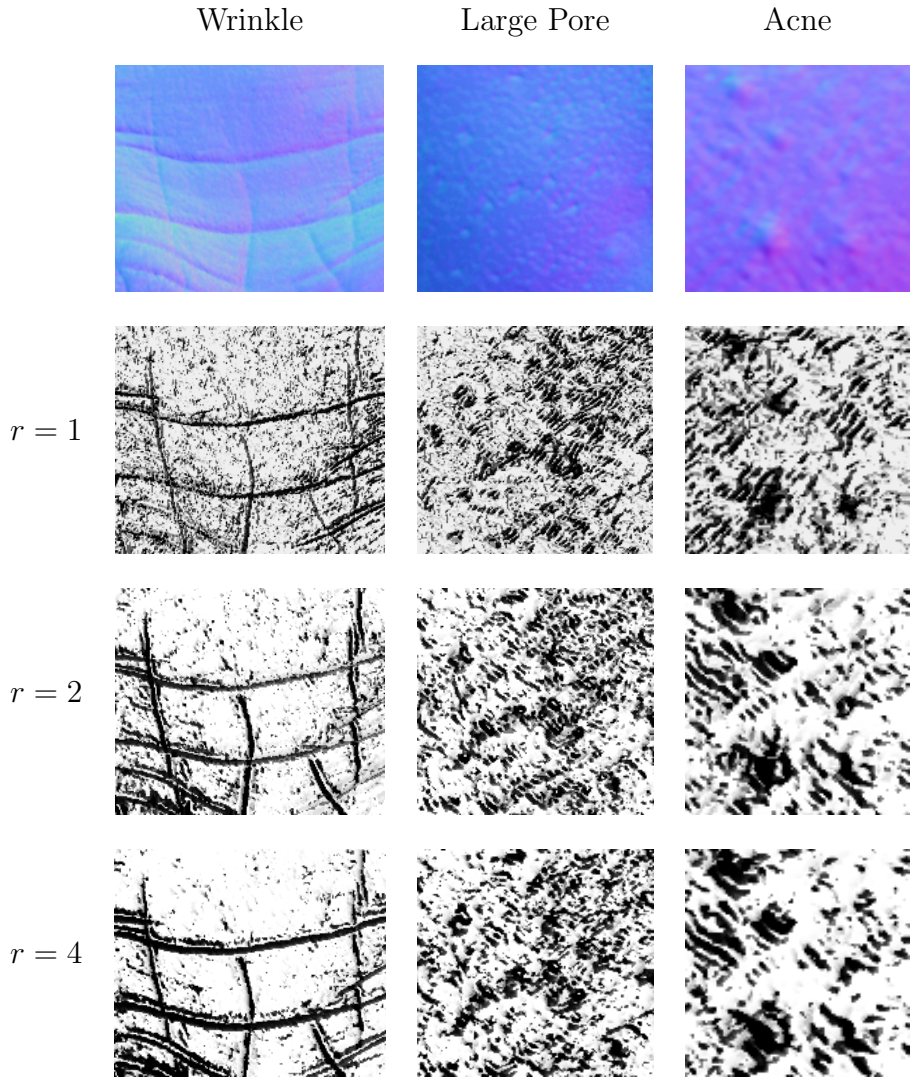


FIGURE 4.8: Local Orientation Pattern of skin normal maps with different radii

4.3.4.2 Second Pattern Function

In the second proposed pattern function, the azimuthal and polar angles of the normal are directly compared. The function has four possible values and is defined by:

$$f_i(\mathcal{N}_0, \mathcal{N}_i) = \begin{cases} 0 & \text{if} \quad \theta_0 < \theta_i \quad \text{and} \quad \phi_0 < \phi_i \\ 1 & \text{if} \quad \theta_0 < \theta_i \quad \text{and} \quad \phi_0 \geq \phi_i \\ 2 & \text{if} \quad \theta_0 \geq \theta_i \quad \text{and} \quad \phi_0 < \phi_i \\ 3 & \text{if} \quad \theta_0 \geq \theta_i \quad \text{and} \quad \phi_0 \geq \phi_i \end{cases} \tag{4.21}$$

where θ_i and ϕ_i are the azimuthal and polar angle of the normal $\mathcal{N}_i$ respectively. Here the required size of the histogram is given by 4^N. This function does not need the extra parameter the previous one had (threshold), although it generates a much bigger feature vector (histogram). While the first function generates (for the standard 8 pixel neighbourhood) a feature vector of length 256, this function generates a 65536 feature vector.

Figure 4.9 shows the Local Orientation Pattern images of three skin patches using the second pattern function with a radius of 1, 2 and 4.

4.3.4.3 Discussion

A glance at the LOP (Local Orientation Patterns) images in Figure 4.8 and Figure 4.9 gives a first idea of the behaviour difference between the two proposed pattern functions. The second pattern function tends to produce LOP images

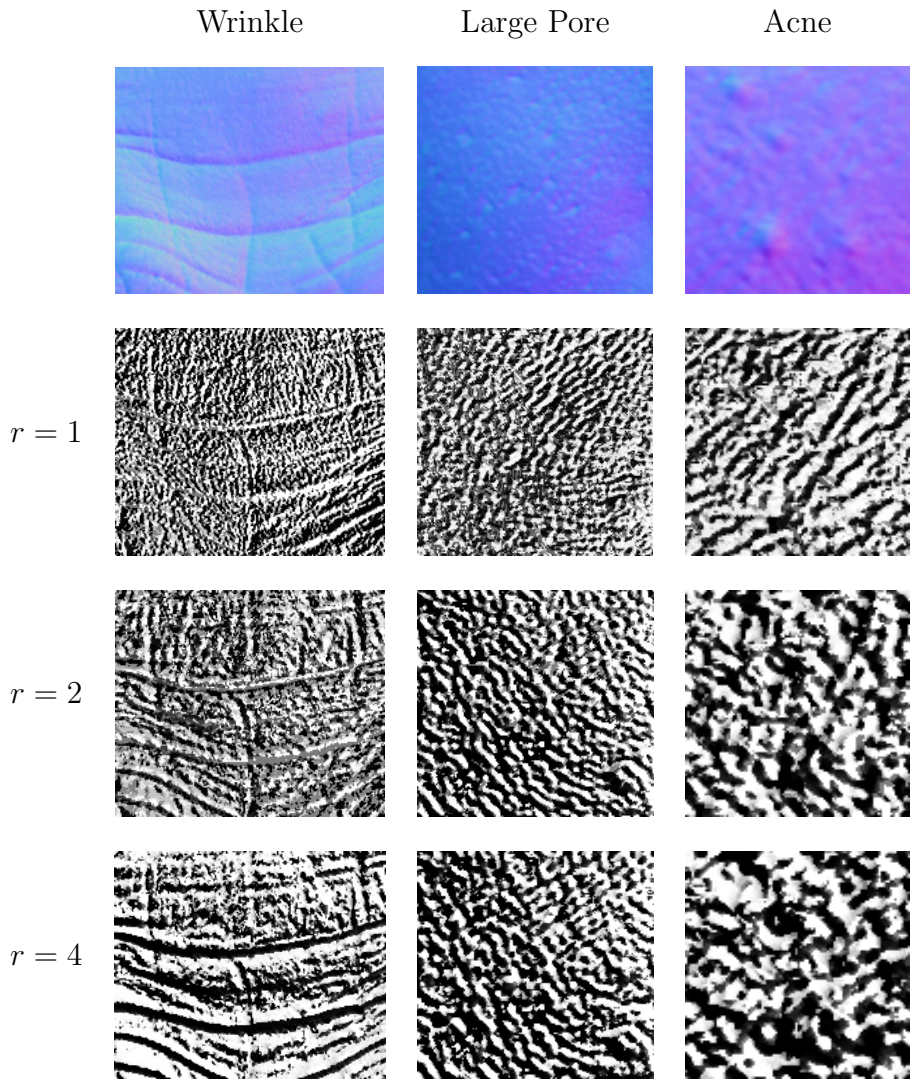


FIGURE 4.9: Local Orientation Pattern of skin normal maps with different radius

with higher frequency. This is probably due to the level of detail generated by using four patterns instead of just two. The important point here is that when using the second pattern function for capturing low frequency properties of a surface, a certain amount of entropy, depending on how fine the surface structure is, can be induced. In our application, we think that it is more appropriate to use it for high frequency skin properties such as pores and some lines and wrinkles, while the first function is more appropriate for capturing lower frequency

conditions as acne.

4.3.5 Classification Results

Table 4.2 gives the classification results using the Local Orientation Patterns as 3D texture descriptors. The results confirm the observation made in the previous (Section 4.3.4.3) about the second pattern function being more appropriate for extracting high frequency conditions as it yields considerably higher performances on classifying pores. However the performances on classifying wrinkle and acne are rather comparable between the two pattern functions.

The overall performances are lower than the ones yielded by Rotation fields pyramid (Section 4.2) which could be due to the fact that the Rotation fields pyramid uses a multi-scale scheme while the Local orientation patterns don't.

TABLE 4.2: Classification results

		Wrinkles			Pores			Acne		
Features \ Sample Size		20	50	80	20	50	80	20	50	80
L.O.P	1st P.F.	.71	.70	.76	.63	.66	.63	.75	.79	.81
	2nd P.F.	.72	.72	.77	.79	.81	.80	.73	.78	.83

4.3.6 Conclusion

We presented in this section a method of characterising 3D surface texture from normal maps. The approach is based on the notion of Texture Spectrum and can be used in any application dealing with 3D surface texture analysis. We introduced two ways of evaluating the normals orientation. The first applies a

threshold on their dot product, while the second variant compares their polar and elevation angles directly. In Chapter 6 we present results using the method for classifying facial skin conditions such as wrinkles, large pores and acne.

4.4 Proposed Method III: Multi-scale Azimuthal Projection Distance

In this section, we investigate extensions of the Azimuthal Projection Distance Image (APDI) introduced in [72] as a 3D surface descriptor for facial Action Unit detection. In their work, the authors used the APDI in the coarse scale and extracted facial macro-structure. However, while these facial macro-structures are adequate for discriminating Action Units, they could not hold enough surface fine-scale deformations to perfectly characterise the skin conditions we are interested in (wrinkles, large pores and acne). The APDI is extended with three main additions:

- Using the local surface normal means instead of a fixed surface mean as a reference for the azimuthal projection.
- The APDI formula is modified to take into account the surface normal azimuthal orientation which is not considered in the original formulation.
- A Multi-resolution analysis scheme is introduced in order to capture different scales of skin deformations.

Each of these extensions are presented in more details in the Sections 4.4.2 and 4.4.3.

4.4.1 Background

Here APDI is briefly recapped, a more detailed presentation being given in Section 2.3.4. For recall, the APDI is a 2D image where the pixels are the projections of the surface normals onto the tangent plane. Given a surface normal at a pixel (i, j), the azimuthal projection is given by:

$$x_{i,j} = k' \cos\theta_{i,j} \sin\left(\phi_{i,j} - \bar{\phi} i, j\right) \tag{4.22}$$

$$y_{i,j} = k' \left(\cos\bar{\theta}_{i,j} \sin\phi_{i,j} - \sin\bar{\theta}_{i,j} \cos\theta_{i,j} \cos\left(\phi_{i,j} - \bar{\phi}_{i,j}\right)\right) \tag{4.23}$$

where $\theta_{i,j}$ and $\phi_{i,j}$ are the polar and azimuthal angles of the surface normal respectively. $\bar{\theta}_{i,j}$ and $\bar{\phi}_{i,j}$ are the polar and azimuthal angles of the mean surface normal over a fixed neighbourhood around (i, j) respectively. Finally $k' = \frac{c}{sin(c)}$ with $c = \sin\bar{\theta}_{i,j} \sin\theta_{i,j} + \cos\theta_{i,j} \cos\left(\phi_{i,j} - \bar{\phi}_{i,j}\right)$

[72] fixed a constant mean surface normal $(0, 0, 1)$ (z-axis direction) which leads to $\bar{\theta} = \frac{\pi}{2}$, $\bar{\phi} = 0$ and $c = \sin\theta_{i,j}$. Thus Equations 4.22 and 4.23 become:

$$x_{i,j} = k' \cos\theta_{i,j} \sin\phi_{i,j} \tag{4.24}$$

$$y_{i,j} = k' \cos\theta_{i,j} \cos\phi_{i,j} \tag{4.25}$$

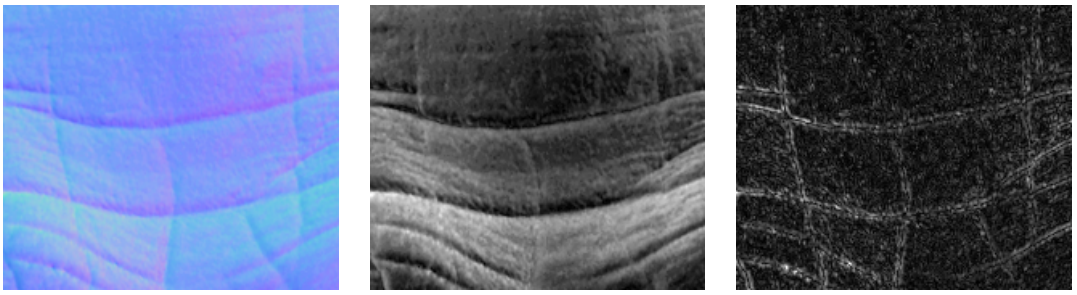
Each pixel value of the Azimuthal Projection Distance Image (APDI) is given by the L^2-norm of $(x_{i,j}, y_{i,j})$:

$$APDI_{i,j} = \sqrt{x_{i,j}^2 + y_{i,j}^2} \tag{4.26}$$

4.4.2 Modified APDI

As stated above, in the original formulation, the authors set a constant surface normal mean ($[0, 0, 1]$) over the whole face, thus projecting about a constant vector across the face. A direct consequence of this is the presence of considerable low frequency information on the APDI, as the mean surface normal constitute the reference about which the normals are projected (the tangent plane that the normal are projected onto is the plane orthogonal to the mean surface normal). While this is suitable for coarse features as facial Action Units, it would introduce notable low frequency bias to the fine skin structures we are interested in. Thus we compute at each pixel a local mean surface normal over a local neighbourhood and use it as projection reference. Hence, in this work we use Equations 4.22 and 4.23 instead of the simplified versions of [72]. Equations 4.24 and 4.25. Figure 4.10 shows example output from the original and our proposed modified

APDI. On the output image of the original APDI, the low frequency is still noticeable whereas in the modified version only the high frequency information is kept.



Input Original APDI Modified APDI

FIGURE 4.10: Example of output from the original and modified APDI

Another modification we made to the original APDI formulation is the introduction of the azimuthal orientation of the surface normal in Equation 4.26 which takes into account only the polar orientation. We illustrate this in Figure 4.11 where we assume that the mean surface normal is aligned to the $z-$axis. It is easy to see that the distance form the centre of projection (red line), which correspond to the original formula, stays constant for all normals with the same polar angle ϕ even though the azimuthal angle θ varies. We overcome this by changing Equation 4.26 to:

$$APDI^m_{i,j} = arctan_2\left(x_{i,j}, y_{i,j}\right)\sqrt{x^2_{i,j} + y^2_{i,j}} \tag{4.27}$$

This corresponds to the arc (blue line) in the projection plane going from the $x-$axis to the projected point and varies with ϕ as well as θ.

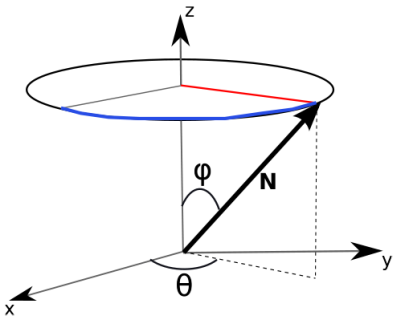
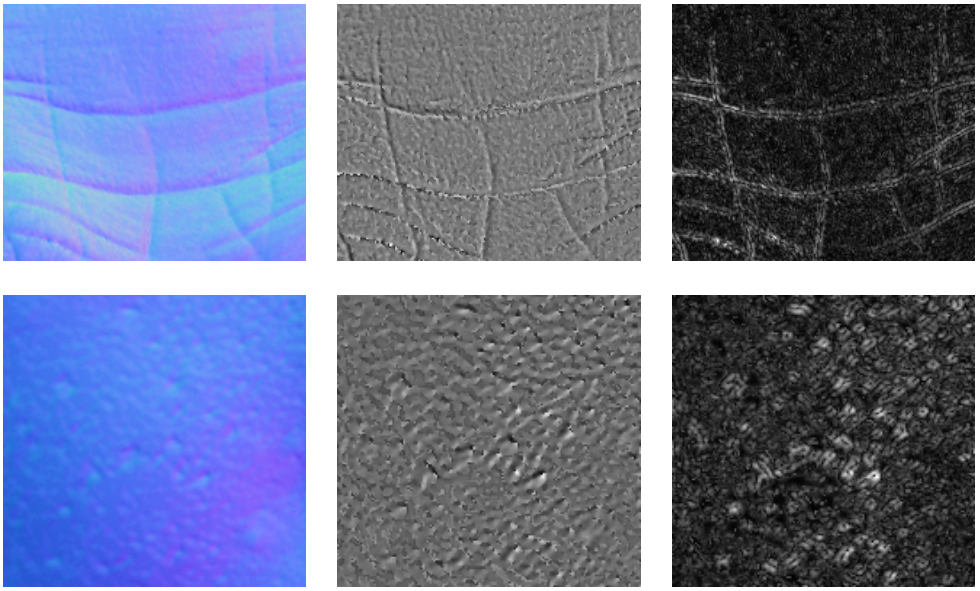


FIGURE 4.11: The distance from the centre of projection (red) is the same for all the normals with the same polar angle, while the arc (blue) from the $x-$axis varies with both the azimuthal and polar angles

Figure 4.12 shows the difference between using the L^2-norm (distance from the centre of projection) or the arc from the $x-$axis. In the first case (L^2-norm), the APDI appears less contrasted in comparison with the second case (arc) which presents more disparity and hence will be more discriminative as shown in the classification results in Chapter 6.



Input With Norm With Arc

FIGURE 4.12: Example of output from considering the L^2-norm (original) and the arc (modified)

4.4.3 Multi-resolution Scheme

We employ a multi-scale scheme of APDIs for analysing the 3D skin texture from dense surface orientations. For a given normal map, a multi-scale APDI pyramid is built by computing the normal map's APDIs at different resolutions. This involves scaling (down or up sampling) of the normal map. Since the surface normals do not satisfy the linearity condition required by classical convolution methods (any linear combination of two normals violates the unit-length constraint), we propose using a geodesic-based normal map scaling algorithm.

4.4.3.1 Scaling Normal Maps

We use Riemannian differential geometry elements to introduce a new metric (geodesic distance) which will allow us to perform linear operations on the normals. We assume the normals to be on a Riemannian manifold and compute all linear operations on a tangent plane that we chose to be constant for all the normals (this is studied in [64]). Let exp_μ and log_μ be the Riemannian Exponential and Logarithm operations with μ as projection axis, the linear combination of N normals $(n_i)_{1\leq i\leq N}$ with coefficients $(\alpha_i)_{1\leq i\leq N}$ can be computed as:

$$f(n_i, \alpha_i) = \exp_\mu \left(\sum_{i=1}^{N} \alpha_i \times \log_\mu (n_i) \right) \tag{4.28}$$

with,

$$\log_{\mu}(n_{\theta,\tau}) : \begin{cases} x = k\cos\theta\sin(\tau - \tau_0) \\ y = k(\cos\theta_0\sin\tau - \sin\theta_0\cos\theta\cos(\tau - \tau_0)) \end{cases}$$

,

$$\exp_{\mu}(t_{x,y}) : \begin{cases} \theta = \sin^{-1}(\cos c\sin\theta_0 + \frac{y\sin c\cos\theta_0}{c} \\ \tau = \tau_0 + \tan^{-1}\frac{x\sin c}{c\cos\theta_0\cos c - y\sin\theta_0\sin c} \end{cases}$$

,

$$c = \sqrt{x^2 + y^2}$$

and

$$k = \frac{c}{\sin c}$$

By definition of the Exponential mapping, the result will always be a unit vector. We base our scaling algorithm on Equation 4.28. As we are only interested in down-sampling, we present an overview of the down-sampling algorithm here in Algorithm 1. The full implementation includes border checking and index checking which we have omitted here for brevity.

We tested the proposed method by comparing a normal map with the result of down-sampling and up-sampling it back. The geodesic method achieves 0.027 mean angular error, while using a classical sampling method on each channel and renormalising back the result gives a mean angular error of 0.183.

Algorithm 1: Normal Map Down-sampling Algorithm

Data: N: normal map, S: scale factor, $[u, v]$: window size
Result: N': Down-sampled normal map

$n_w \longleftarrow \frac{width(N)}{S}$
$n_h \longleftarrow \frac{height(N)}{S}$
for $i \longleftarrow 1$ **to** n_w **do**
  for $j \longleftarrow 1$ **to** n_h **do**
    $Tmp \longleftarrow 0$
    for $w \longleftarrow i - \frac{u}{2}$, $k \longleftarrow 1$ **to** u **do**
      for $y \longleftarrow j - \frac{v}{2}$, $l \longleftarrow 1$ **to** v **do**
        $Tmp \longleftarrow Tmp + Log_{\mu}\left(N\left(w, y\right)\right)$ $y \longleftarrow y + 1$
      end
      $w \longleftarrow w + 1$
    end
    $N'(i, j) \longleftarrow Exp_{mu}\left(\frac{Tmp}{u \times v}\right)$
  end
end

4.4.3.2 Multi-Resolution APDI

To characterise the 3D skin texture, we build a multi-resolution pyramid of APDIs by down-sampling the normal map to different levels. At each level, the APDI is re-computed from the corresponding down-sampled normal. The high levels contains higher frequency details adequate for texture analysis. The more we go down to the lower levels, the more we lose high frequency details, but the low frequency changes related to the overall shape are highlighted.

Figure 4.13 shows examples of image output of the modified multi-resolution APDI for 3 skin patches with presence of wrinkle, large pores and acne respectively. It is interesting to notice how, at different scales, the level of high frequency information that is captured changes. For example, considering the patch with acne, one can see that on the first level, only the fine skin structure is captured. It is clear that stopping the texture extraction at that level would capture

only partial information about the skin disruption and would certainly miss the big skin spots. These are better captured in the subsequent levels as shown Figure 4.13.

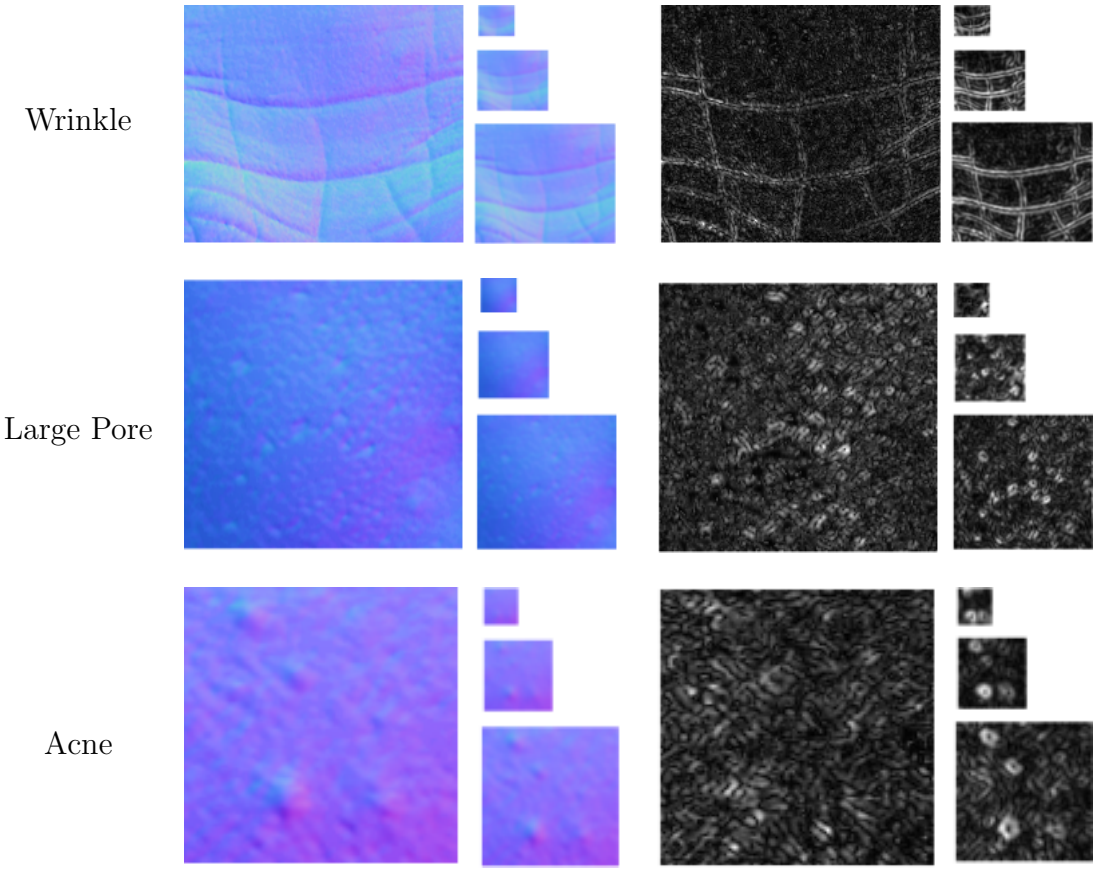


FIGURE 4.13: 4−level APDI pyramid for 3 skin patches

4.4.4 Feature Extraction

To extract features from a given normal map patch, the multi-resolution APDI pyramid is built. Then a grey level histogram is computed at each level of the pyramid and concatenated together. This produces a relatively large feature vector depending on the number of levels and the histogram resolution (e.g. number of bins). For example an 128-bins histogram with a 4-level pyramid

will produce a feature vector of length 512. This can be reduced using feature selection techniques as we present in Chapter 6.

4.4.5 Classification Results

The classification performances using the original formulation of the APDI and our modified version are presented in Table 4.3. The results show a clear improvement of the modified APDI over the original version. The M. APDI with a depth of 1 differ from the O. APDI in the choice of the projection reference and the consideration of the azimuthal orientation, yet it outperforms the O. APDI even without the multi-scale scheme. Introducing a multi-scale analysis improves even more the performances of the M.APDI. With a multi-scale depth of 4, the increase in classification performances goes up to 0.16.

TABLE 4.3: Classification results

		Wrinkles			Pores			Acne		
Features \ Sample Size		20	50	80	20	50	80	20	50	80
O. APDI	-	.62	.61	.65	.63	.60	.62	.60	.63	.63
M-APDI	$depth = 1$	.62	.65	.68	.62	.62	.64	.61	.65	.64
	$depth = 2$	.69	.70	.73	.71	.70	.72	.68	.67	.70
	$depth = 4$	.75	.78	.81	.74	.76	.73	.72	.74	.75

4.4.6 Conclusion

This section presented another approach to characterising 3D surface texture based on extensions of the Azimuthal Projection Distance Image (APDI) introduced in [72]. We proposed three extensions of the original formulation. These include:

- working with local surface normal means instead of a fixed surface mean as reference for the azimuthal projection.

- modifying the APDI original formula to take into account the surface normal's azimuthal orientation which is not considered in the original formulation.

- introducing a multi-resolution analysis scheme in order to capture different scales of skin deformations.

We show in Chapter 6 that with these extensions, the modified multi-resolution APDI considerably outperforms the original APDI in classifying the skin condition we are interested in (wrinkle, large pores and acne).

4.5 Summary

This chapter has described the different contributions we propose for 3D surface texture representation. We first presented a comparative study where 2D texture measures are compared against simplistic 3D extensions. As shown in Chapter 6, the 3D descriptors outperform all their 2D counterparts in classifying facial skin condition such as wrinkles, large pores and acne. We then presented in detail the three proposed 3D surface texture descriptors proposed in this thesis. These are:

- **Rotation Fields** where 3D surface texture information is extracted by separating high frequency data from the low frequency in the form of rotation map.

- **Local Orientation Patterns** inspired by the notion of Texture Spectrum where at each pixel, the surface normal is compared against the ones surrounding it.

- **Multi-scale APDI** which is an extension of the Azimuthal Projection Distance Image where the normal are projected onto the tangent space before applying common texture measures such as LBPs or Gabor filtering.

All three descriptors were evaluated on classifying facial skin condition (wrinkles, large pores and acne) on an extensive dataset collected in the context of this thesis. The next chapter presents in detail the data collection process and Chapter 6 gives and discusses the results of the different experiments conducted to evaluate the texture descriptors presented in this thesis.

Chapter 5

Data Collection and Annotation

5.1 Introduction

Chapter 4 introduced various proposed methods of characterising 3D surface texture from facial skin normal maps. To evaluate these different methods, we have collected an extensive dataset of high resolution 3D faces along with reflectance data containing both fine geometric and photometric properties of the facial skin, down to the level of the pores. The dataset consists of 50 subjects with an age range going from 19 to 68 and presenting various skin types. We have used a Lightstage, which is a dome of controlled lights, to capture both the 3D models and skin reflectance.

To annotate the collected data, we have run an experiment where 25 participants have been asked to rate skin patches from different facial regions according to the presence/visibility of wrinkles, pores and acne.

We give here a detailed presentation of the different steps involved in the data collection process.

5.2 Existing Datasets

A considerable number of 3D face datasets have been proposed over the last few years. While most of these aim to provide data for face recognition or expression recognition task, only a few offer resolutions appropriate for fine texture analysis. For example, the BU-3DFE (Binghamton University 3D Facial Expression) [94] captured geometric meshes at a resolution of 35000 vertices. The Bosphorus dataset [73] collected high resolution texture (1600×1200 pixels) but the 3D meshes are too coarse for extracting fine 3D surface texture information. The EFPRI (ETH Face Pose Range Image) database [3] captures face depth map at only 150×200 pixels.

The photometric stereo method, introduced in [93], which allows us to obtain surface orientation from multiple images with varying illumination, is one of the most appropriate ways of capturing 3D surface fine texture as the resolution of the final data is only limited by the sensor used. While one of the main issues of photometric stereo is aligning the different images to compensate for motion of the target, powerful optical flow algorithms are available to tackle this [4–6]. 3D face datasets that use photometric stereo include the Photoface database [95] and the 3D Relightable Facial Expression (ICT-3DRFE) database [79]. While the first was captured with low cost cameras, the latter is captured using a

lightstage which is to date the most advanced setup for capturing surfaces' fine structure. Despite providing highly detailed 3D data, the ICT-3DRFE database is not suitable for this work as the age range and skin type covered by the dataset is limited. We collected a new dataset covering a wider age range and skin type using a lightstage. The next sections present our capture setup and the data collection process in details.

5.3 The Lightstage

The lightstage is a 3D surface acquisition device first proposed in [53]. It is a dome of controlled LEDs that captures surface geometry along with specular and diffuse reflectance data. The Lightstage is based on spherical gradient photometric stereo techniques. It operates by capturing the target under different spherical gradient illumination patterns and uses polarisation techniques to separate diffuse from specular reflectance. The final output data consist in dense surface normal fields whose resolutions are limited only by the resolution of the sensor used to capture the gradient images.

5.3.1 Our Lightstage Setup

In collaboration with the University of York, we have developed a Lightstage of 1.58m diameter with 41 controllable LEDs evenly placed on the dome's frame (Figure 5.1) . Linear polarisers are mounted on every LED and on the photometric camera which is a Nikon D200. The polariser mounted on the photometric

camera is equipped with an opto-electric system that allows us to automatically flip the polarising direction and capture cross-polarised data (this is used for specular and diffuse data separation). In addition to the photometric camera, six Canon 7D DSLRs are used for multiview images to reconstruct a base mesh. All these components (LEDs, cameras and opto-electric polariser) are centrally controlled via a programmable Mbed board [1].

FIGURE 5.1: The Lightstage used for our data collection

5.3.2 Processing the Raw Data

Our system uses a multi-pose photometric acquisition, which allows us to capture facial reflectance and micro-structure data from ear to ear. We consider three poses: frontal, left profile and right profile. For each pose, the raw output data consists of two sets of images:

[1] https://www.mbed.com

- **14 spherical gradient illumination images**, half of which are cross-polarised in order to measure only the diffuse component of the reflectance (the specular component is obtained by subtracting from the seven other non filtered images the corresponding diffuse reflectance). In each set (7 filtered and 7 non-filtered), the subject is captured under 6 gradient illumination patterns ($X-$gradient, $Y-$gradient, $Z-$gradient, $\bar{X}-$gradient, $\bar{Y}-$gradient, $\bar{Z}-$gradient) and a constant illumination C where all the LEDs are on. This gives a total of 14 images, 7 diffuse (X_d, Y_d, Z_d, $\bar{X}_d$, $\bar{Y}_d$, $\bar{Z}_d$ and C_d) and 7 specular (X_s, Y_s, Z_s, $\bar{X}_s$, $\bar{Y}_s$, $\bar{Z}_s$ and C_s). We show in Figure 5.2 the set of 14 gradient images of the frontal pose in the two upper rows and the calculated specular gradient images in the last row.

- **Mutli-view images** captured at the same time, in the constant illumination, by 6 SLRs positioned to provide enough overlapping coverage of the face. Figure 5.3 shows the six multiview images of the frontal pose.

5.3.2.1 Normal Field Calculation

[53] established the relation between the surface normal and the first three gradient illumination images. They did not consider the complement illuminations. In their formulation, the diffuse and specular surface normals are given by:

$$N_d = \frac{\left[\left(\frac{X_d}{C_d} - \frac{1}{2}\right), \left(\frac{Y_d}{C_d} - \frac{1}{2}\right), \left(\frac{Z_d}{C_d} - \frac{1}{2}\right)\right]}{\sqrt{\left(\frac{X_d}{C_d} - \frac{1}{2}\right)^2 + \left(\frac{Y_d}{C_d} - \frac{1}{2}\right)^2 + \left(\frac{Z_d}{C_d} - \frac{1}{2}\right)^2}} \tag{5.1}$$

FIGURE 5.2: Spherical gradient illumination images of the frontal pose. 1st row: diffuse only. 2nd row: diffuse & specular. 3rd row: specular only (obtained by subtracting the diffuse images from the diffuse & specular)

FIGURE 5.3: The multiview set of the frontal pose

and

$$N_s = \frac{1}{\bar{N}}(u + v) \tag{5.2}$$

with v being the view vector, $(0, 0, 1)$ in our case, and

$$u = \frac{\left[\left(X_s - \frac{C_s}{2}\right), \left(Y_s - \frac{C_s}{2}\right), \left(Z_s - \frac{C_s}{2}\right)\right]}{\sqrt{\left(X_s - \frac{C_s}{2}\right)^2 + \left(Y_s - \frac{C_s}{2}\right)^2 + \left(Z_s - \frac{C_s}{2}\right)^2}} \tag{5.3}$$

The complement gradient images were introduced by Wilson et al. who proposed a more optimal way of calculating the surface normals [91]. The complement gradient images $\bar{X}$, $\bar{Y}$ and $\bar{Z}$ are such that:

$$X + \bar{X} = Y + \bar{Y} = Z + \bar{Z} = C \tag{5.4}$$

Combining Equations 5.1, 5.2 and 5.4, they obtained for both specular and diffuse normals:

$$N_{d(s)} = \frac{\left[\left(X_{d(s)} - \bar{X}_{d(s)}\right), \left(Y_{d(s)} - \bar{Y}_{d(s)}\right), \left(Z_{d(s)} - \bar{Z}_{d(s)}\right)\right]}{\sqrt{\left(X_{d(s)} - \bar{X}_{d(s)}\right)^2 + \left(Y_{d(s)} - \bar{Y}_{d(s)}\right)^2 + \left(Z_{d(s)} - \bar{Z}_{d(s)}\right)^2}} \tag{5.5}$$

The Wilson formulation [91] is proven to be more optimal than the original method proposed by Ma et al. [53]. This is because it is more robust to gradient image quality as the darks regions in one gradient image are likely to be well lit in the complement image. Also gradient images and their corresponding complement tend to balance out possible ambient shadow effects.

As the Wilson method appears to improve the calculated normal quality, all our normal fields are generated using it. We show in Figure 5.4 an example of the specular and diffuse normal fields obtained.

FIGURE 5.4: Diffuse (top) and specular(bottom) normal fields for the 3 poses (left profile, frontal and right profile)

5.3.2.2 Geometry Recovery

We use Agisoft Photoscan [2], a commercial Shape From Motion (SFM) system, to recover the geometry from 21 multi-view images ($(6+1) \times 3$ images from the six multi-view cameras and the photometric camera at constant illumination for each of the three poses). The advantage of using Shape From Motion is that no prior geometrical calibration is required. The camera position and rotation are estimated at the same time as the scene structure.

Although the meshes produced using SFM techniques may not be of very high quality, they generally hold enough low frequency information to be considered

[2] http://www.agisoft.com/

as a rough estimation of the coarse shape. In order to improve the geometry quality, we use the previously computed normal fields and the mesh generated by Photoscan as base geometry in an iterative refinement process proposed in [60]. The main idea is to use the accurate high frequency information present in the measured normal fields to refine the base mesh and inversely use the mesh coarse shape to correct the low frequency bias present in the normal fields. Figure 5.5 shows an example of refinement results.

5.3.2.3 Pose merging and texturing

As mentioned above, in order to have an ear-to-ear coverage of the face, we capture it in three different poses: frontal, left profile and right profile. We merge the photometric data (albedo and normal maps) from the three poses following these three steps (a more detailed description of the approach is given in [14]):

- **Image sampling:** the 3D mesh and the camera parameters are used to determine the set of visible vertices for each view. Occlusions are tested using a z-buffer.
- **Mesh segmentation:** the 3D mesh is uniformly segmented using a farthest-point strategy. This results in a set of overlapping patches having the same size.

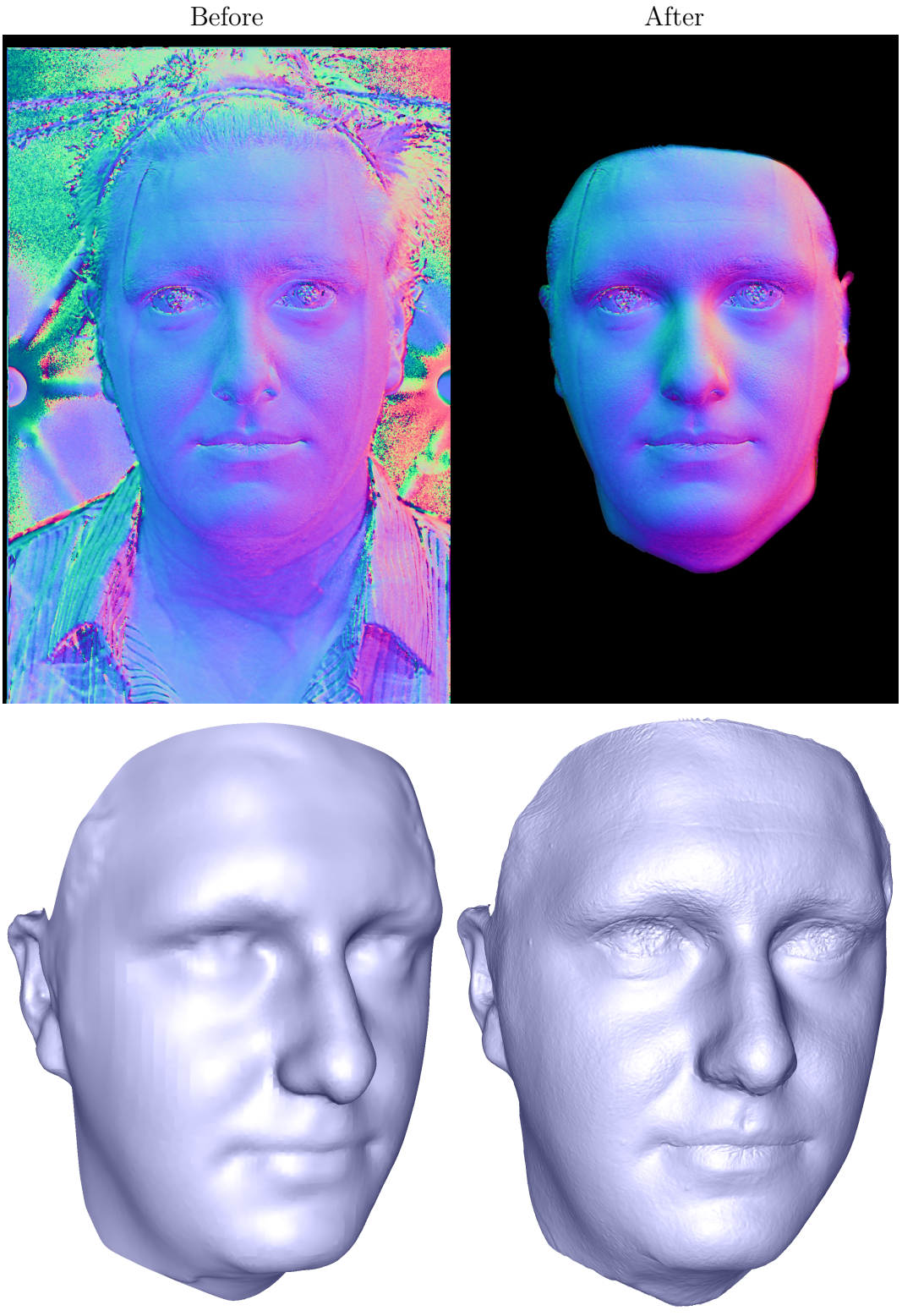


FIGURE 5.5: Result of normal and mesh refinement

- **Stitching:** Using the image sampling results, the best view is voted for each patch. Then the patch's projections in the UV space are stitched in

the gradient domain to produce the final texture.

Figure 5.6 shows an example of the different photometric data in the three poses and the result of their stitching.

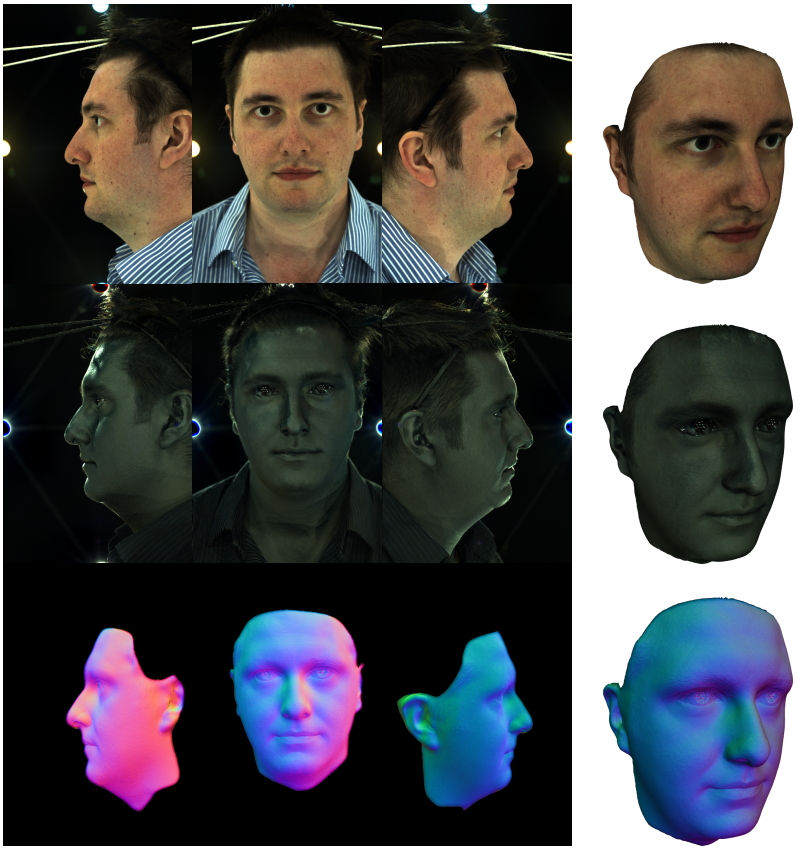


FIGURE 5.6: Pose Merging

5.3.2.4 Region Segmentation

Each face is segmented into 14 regions using a 3D template (set of landmarks) manually adjusted on the face (Figure 5.7). As all processing (analysis or synthesis) is done on the measured normal maps, this segmentation is projected on the 2D texture space of each of the three photometric poses using the corresponding camera parameters.

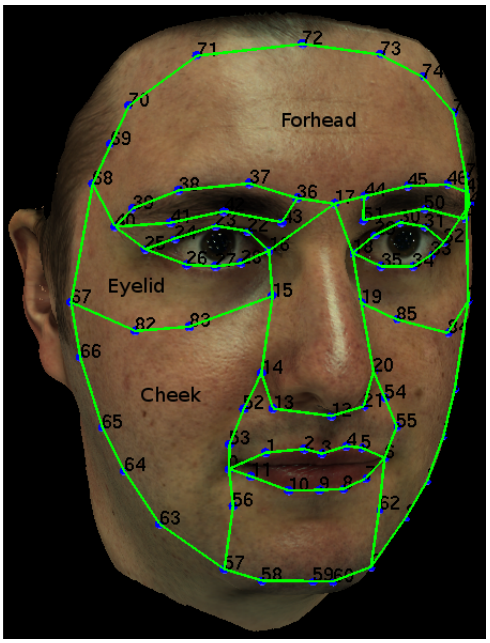


FIGURE 5.7: Template used to segment the captured face in regions of interest

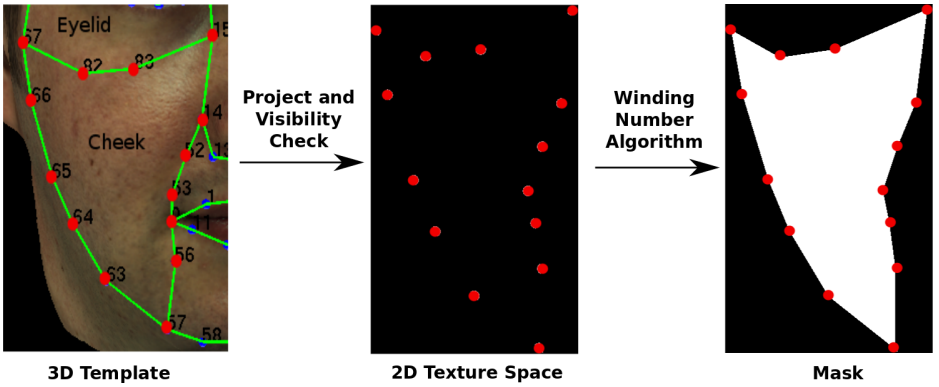


FIGURE 5.8: Example of region mask construction (left cheek on frontal pose)

For each region, we first project its corresponding landmarks onto each of the poses using the camera parameters and a visibility calculation. This results in a set of 2D points in the texture space on each pose. Then we use a winding number algorithm [38] to compute the polygon formed by this set of points for each pose. This polygon is used as a mask for the corresponding region on a given pose. Figure 5.8 shows an example of region mask construction of the left cheek on the frontal pose.

5.4 Data Collection and Rating

5.4.1 Data Collection

We captured a total of 50 faces presenting various skin conditions (wrinkles, acne, visible pores). Figure 5.9 gives an overview of the different skin conditions present in our dataset. The subjects ages ranged from 19 to 68 years old. In addition to the facial images, various extra information about the subjects (age, sex, height, weight, eye color, hair color, makeup, ethnic origin, etc.) were collected. Male participants were more represented than female, with 41 men for 9 women. Various ethnic groups were represented as well, although the majority were Caucasian.

5.4.2 Data Annotation

For data annotation, we conducted a psychophysical experiment where a number of human participants were presented with skin patches from different regions of the face and asked to rate them on a scale of 1 to 5 according to the presence and visibility of wrinkles, acne and pores. The participants were mainly students and staff from the Computer Science department with no training on assessing human skin condition.

We considered only three regions of interest which were: cheek, forehead and eye-corner, as these are the regions where most of the skin conditions we are interested in occur. All the faces were segmented using the generic template shown

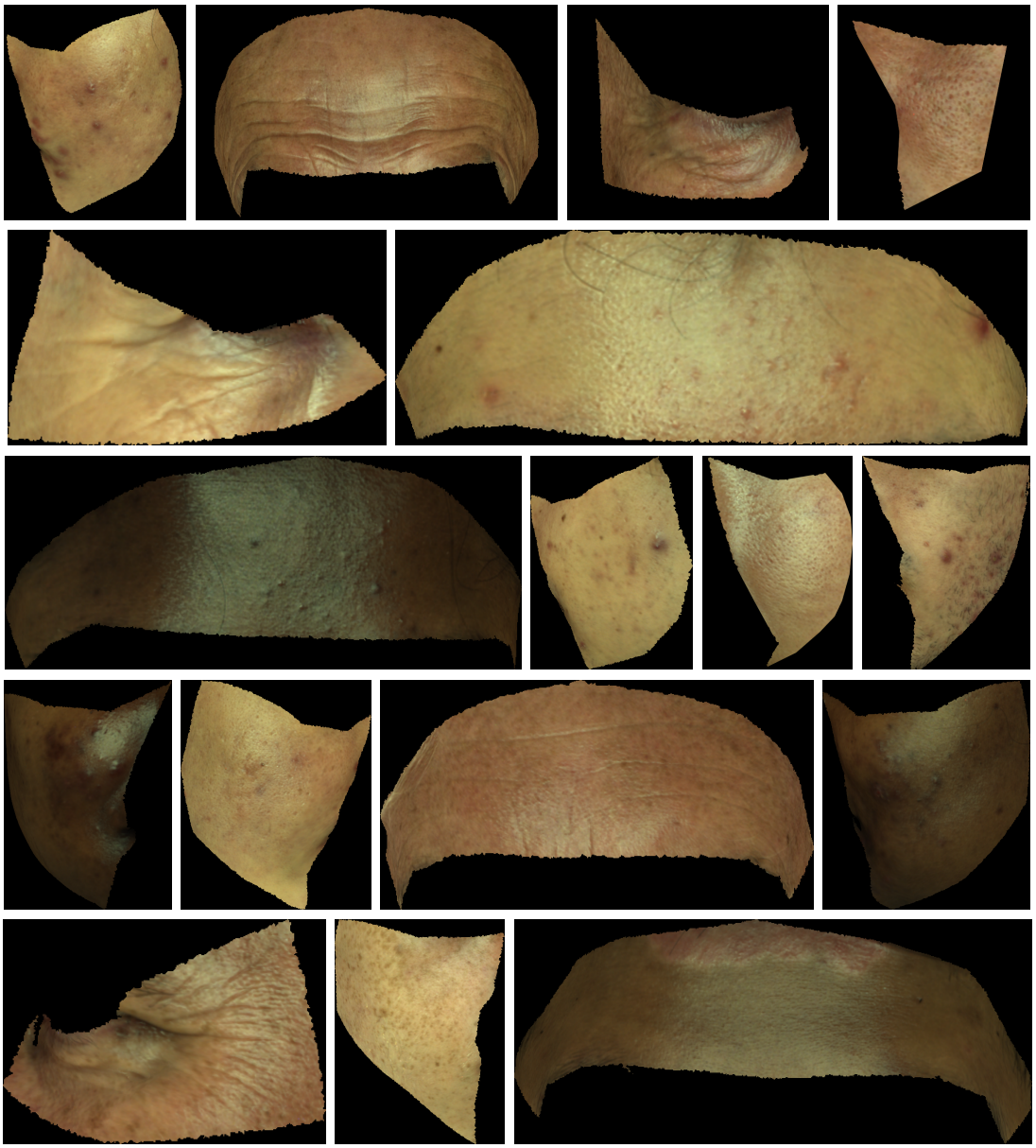


FIGURE 5.9: Overview of the different skin conditions present in our dataset

in Figure 5.7. A photo-realistic animation is made for each patch showing it in different angles with a fixed point light. The photo-realism of this animation is critical to the rating process as the apparent texture of the skin is strongly affected by the lighting and viewing conditions the patch is rendered with. Figure 5.10 shows two skin patches rendered in two different angle-of-views and the difference in apparent texture is clearly noticeable.

Our rating platform was set as a web application (Figure 5.11). The pre-rendered animation of each skin patch is played to the participant at least once before any rating can be entered. The participant has the possibility to re-run the animation as many times as they wish or change between angle-of-view using a slider control. To reduce potential bias in ratings introduced by the order in which the patches are displayed, the sequence allocated to each participant is randomised.

We assume that most of the skin conditions we are interested in are more or less symmetrical across the face (i.e. if a subject presents acne or large pores on the left cheek, it is more likely that the same condition will be found on the right cheek). Thus for each subject, instead of presenting both the left and right cheeks or eye-corner to the raters, only one is randomly picked.

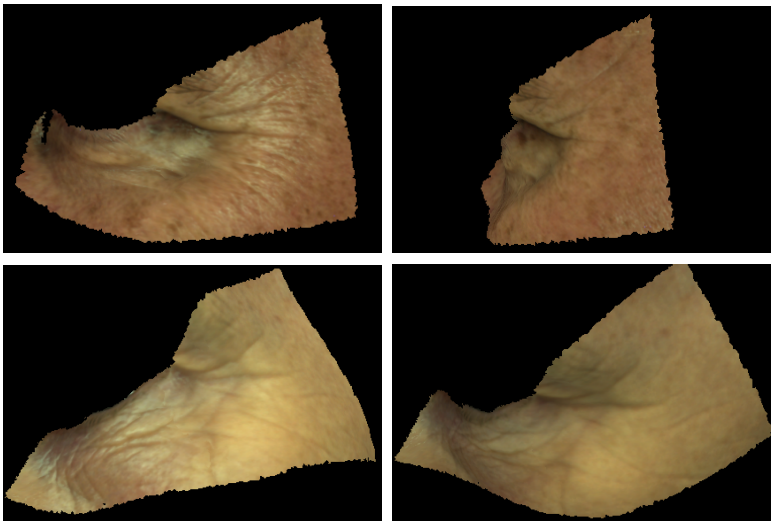


FIGURE 5.10: Change in apparent texture when viewpoint varies

To reduce the rating time and minimise the risk of having participants withdraw before finishing a session, the patches are categorised in blocks according to their location on the face. Thus we have three blocks (cheek, eyelid and forehead) of 50 patches each. A participant choses a block to start with. When they finish

rating one block, they have the possibility to chose another one to rate or stop there.

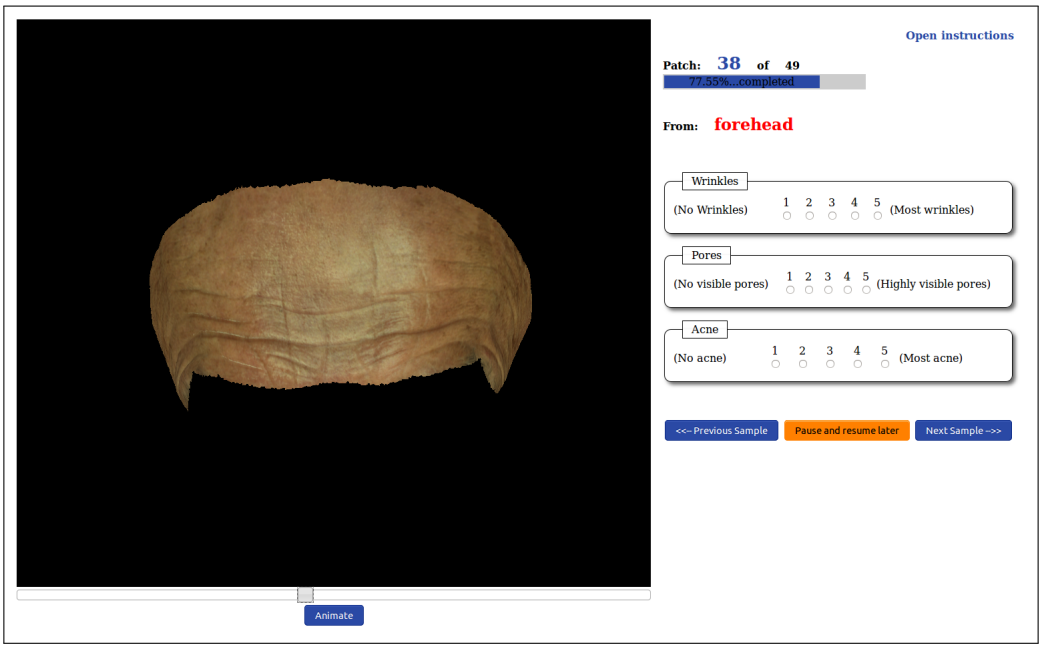


FIGURE 5.11: Rating platform for the psychophysical experiment

Judgement of facial skin texture is a rather subjective task. The way people perceive and quantify the skin conditions that we are interested in will certainly be affected by many factors related to their own personal experience. Therefore, for our dataset to be reliable, it is critical to get it rated by many individuals. This also allows analysis of the correlations between how different people perceive these skin conditions. A total of 25 participants have rated the dataset, with almost all of them having rated at least two blocks.

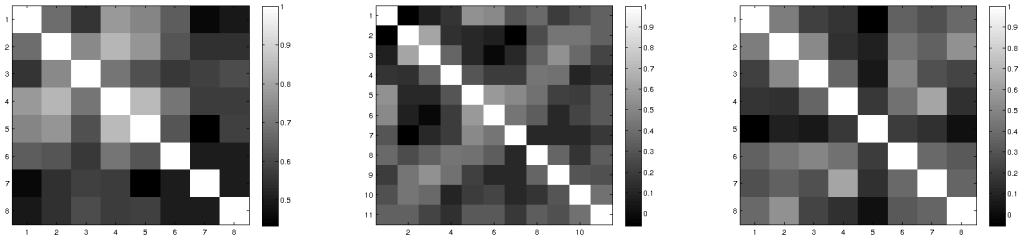
5.4.3 Inter-rater Agreement

As the data is rated by 25 participants, each sample has a set of ratings given by different individuals. Therefore, it is critical for the dataset's reliability to present

a certain level of agreement between ratings provided by different participants. Table 5.1 presents various correlation and agreement measures computed on the raw ratings. Theses show relatively low correlation and agreement between the raters. Apart from the Cheek region, all the different correlation measures are far below 0.5.

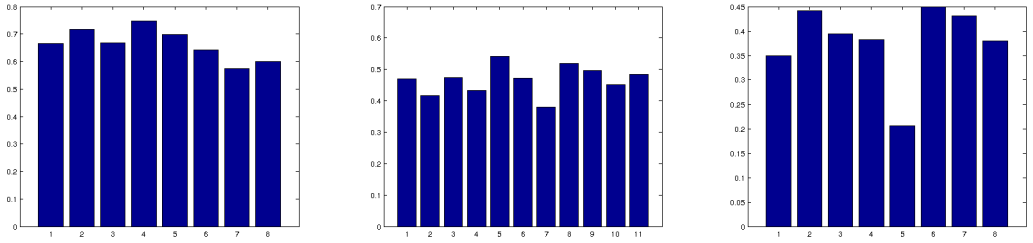
TABLE 5.1: Various Correlation and Agreement Measures on Raw Ratings

Region	Nb.Rater	Correlation			Agreement	
		Pearson	Spearman	Kendal	Fleiss Kappa	Kripp alpha
Cheek	8	0.641	0.639	0.610	0.170	0.403
Eyecorner	11	0.209	0.207	0.200	0.087	0.104
Forehead	8	0.276	0.273	0.260	0.074	0.073



Cheek Eye-corner Forehead

FIGURE 5.12: Correlation between each pair of raters



Cheek Eye-corner Forehead

FIGURE 5.13: Correlation between each rater and all the others

The low correlation measures on the raw data suggests strong disagreement between raters. This can be due to differences in judgement or not genuine ratings. We have computed for each region the correlation between each pair of raters

(Figure 5.12) and the correlation between each user and all the others (Figure 5.13).

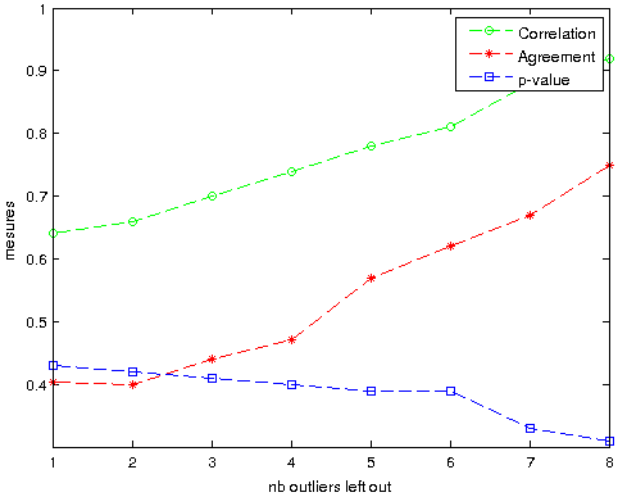
To achieve higher interrater agreement, we experimented with excluding those participants who correlate the least with the rest. The participants are excluded successively by ascending order of correlation to the rest given by the charts in Figure 5.13, starting with the one with the lowest value. Figure 5.14 shows for each region the distribution of correlation and agreement in function of the number of excluded participants. This also shows the effect of the statistical significance (in terms of p-value) of the resultant correlation. Concerning the inter-rater agreement, the exclusion of raters do not introduce substantial increase of agreement, apart from the case of "Cheek" rating. Statistically, correlation and agreement, although both related multivariate data inter-similarity, measure different concepts. While agreement tries to measure how two or more variables are close or the same, correlation measures how similarly they evolve. So two perfectly correlated variables can have very poor agreement. As that is the case for our ratings, we consider the correlation, instead of the inter-rater agreement, as criteria for excluding raters. The exclusion policy we used is as follow: we keep as the maximum number of raters that achieves at least a correlation superior or equal to 0.5.

5.5 Conclusion

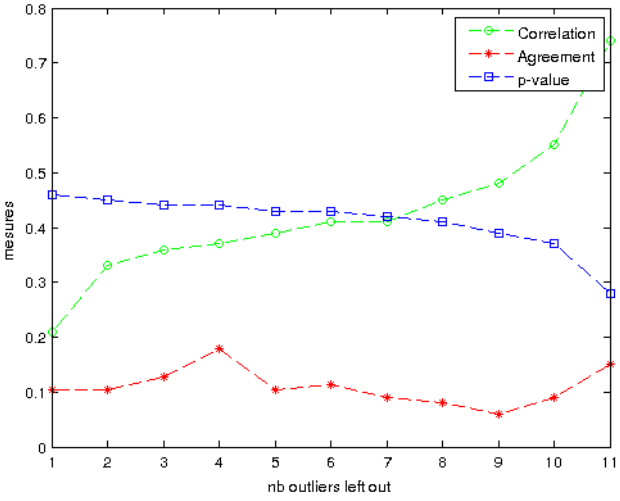
In this section, we presented the different steps undertaken to collect and annotate our 3D face dataset. We have set up a Lightstage which consists of a spherical dome of 1.58m diameter with 41 controllable LEDs and 7 DSLR cameras to capture highly detailed photometric and multi-view data. A cross-polarisation technique is used to separate the specular component of the photometric data from the diffuse component allowing us to independently measure the surface and the subsurface properties of the facial skin.

The capture is realised in three different poses (frontal, left profile and right profile). This allows us to cover the whole face from ear to ear, in contrast to previous datasets such as the 3DRFEE [53]. This is particularly useful to our application as it allows us to take into account all the locations on the face where the skin conditions we are interested in may occur.

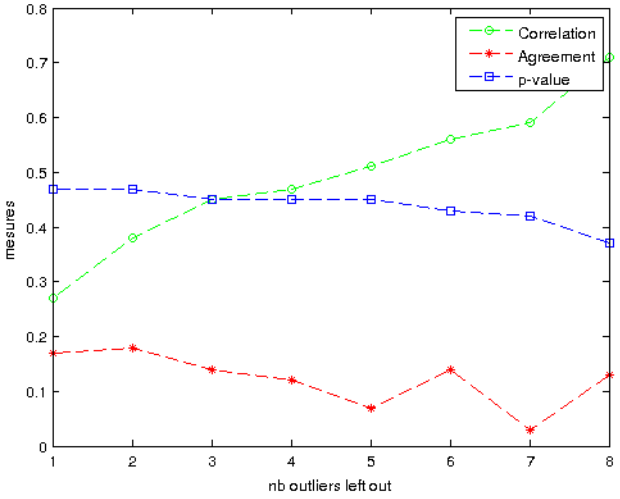
We have collected a total of 50 faces with various skin conditions. Each of these is segmented into regions (forehead, cheek, eye-corner etc.), photo-realistically rendered and rated by different individual in terms of presence of wrinkles, large pores and acne.



Cheek



Eye-corner



Forehead

FIGURE 5.14: Correlation distribution as a function of the number of excluded participants

Chapter 6

Experiments and Results

6.1 Introduction

In the two previous chapters we have presented the 3D surface texture characterisation methods we propose and the different steps involved in our data collection. In this chapter, we will describe the experiments conducted to test the proposed methods on the task of classifying three skin conditions which are wrinkles, large pores and acne present in the collected data.

First, we will give a detailed description of our experimental setup. This includes:

- **a data sampling strategy** that will describe how the samples we will use as Machine Learning inputs are extracted from the original data.
- **a baseline method** which corresponds to the state-of-the art in 3D surface texture representation. For this, we implement a BTF (Bidirectional

Texture Function) based method (as described in Section 6.2.2) that will operate on the photometric specular albedos.

Finally we will give the classification results obtained using each of the methods presented in Chapter 4 and the results yielded by the baseline method.

6.2 Experimental Setup

As described in Chapter 5, the dataset consists of 50 3D face model along with reflectance data (normal map and albedo) in three poses (frontal, left profile and right profile). On the analysis side, as stated in Chapter 4, all the 3D surface texture descriptors proposed operate directly on the specular normal maps whereas the 2D descriptors are extracted from the specular albedo under a direction-independent lighting environment. We also use a BTF Image-Texton based method as baseline, which will use the full set of specular albedo under the 7 illuminations and three poses. This is detailed in Section 6.2.2.

This means that all the texture processing presented in this work will be done in the texture coordinate space. We presented in Section 5.3.2.4 how each of the three regions of interest (forehead, cheek or eye-corner) is projected as a mask on the texture space corresponding to each of the three different poses. However only the forehead is visible in all three poses. The cheeks and eyelids are visible only in the frontal and one profile poses. This yields seven patches for each face, thus 350 patches overall. Figure 6.1 shows the three different poses

of a subject and the masks corresponding to the seven patches in the texture coordinate space.

6.2.1 Dataset Sampling

Due to two issues, each of these 350 patches cannot be considered as individual samples for feature extraction and Machine Learning:

- First, they do not have regular shapes on the edges. This poses a problem when trying to use filters due to the irregular edges. For example, in Figure 6.2, applying a Gabor filter on an eye-corner patch at location **B** will be biased by the *"non-contribution"* of the masked out zone and the result is not theoretically comparable to the result obtained when applying the filter at location **A**.

- Second, their sizes are variable. It would be unfair, without further normalisation, to compare a feature vector extracted from a whole cheek patch to one extracted from an eye-corner patch which is generally more than five times smaller.

To overcome theses two issues, we segment each patch into smaller square patches with a fixed size and the constraint that these should all be totally inside the patch as shown in Figure 6.3. The verification is easily done using a Point-in-polygon algorithm. In our case we used a ray-casting method. The segmentation into square regions obviously introduces another problem which is excluding some

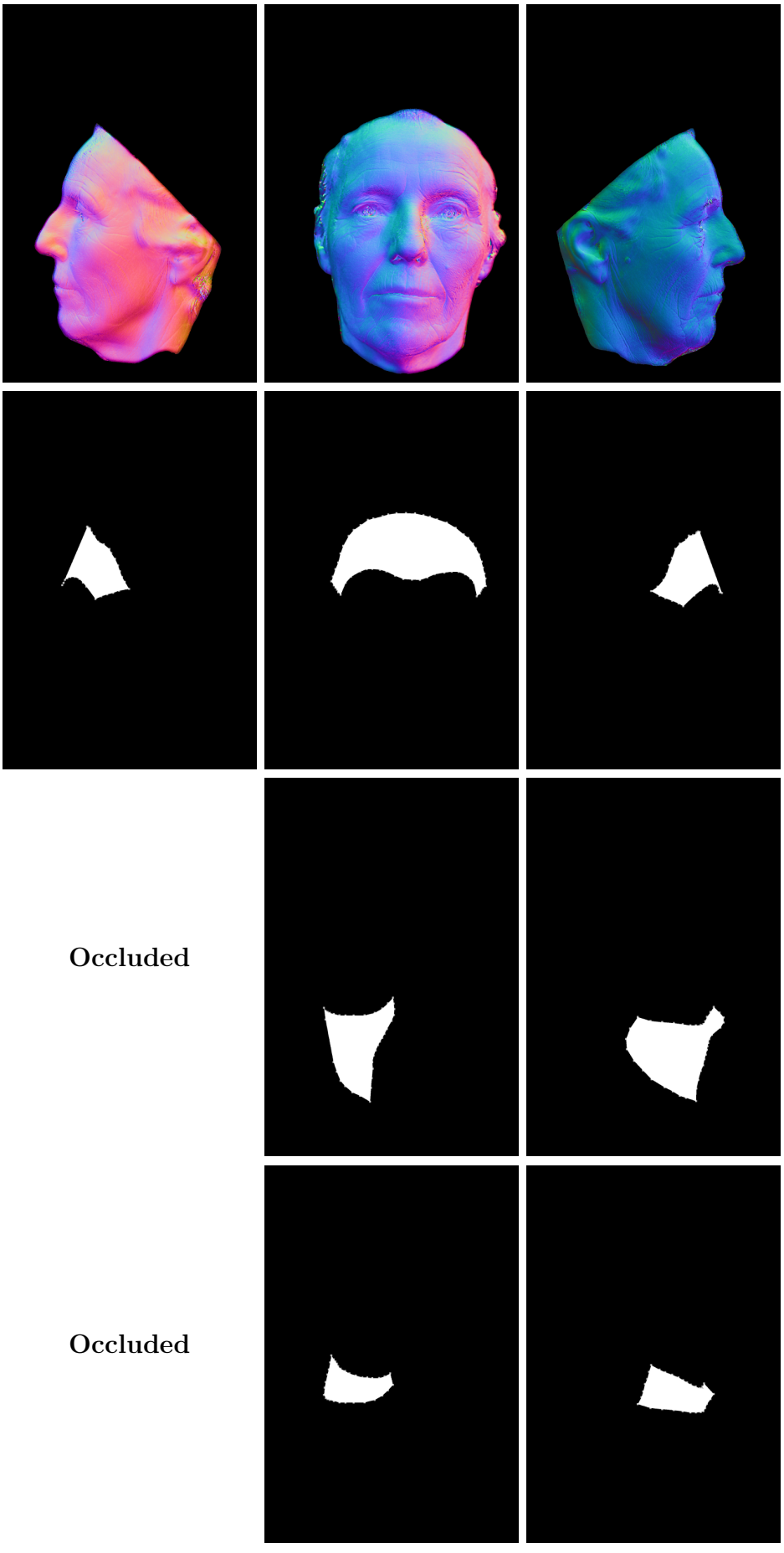


FIGURE 6.1: Regions mask for a subject: forehead (2nd row), right cheek (3rd row) and right eye-corner (4th row)

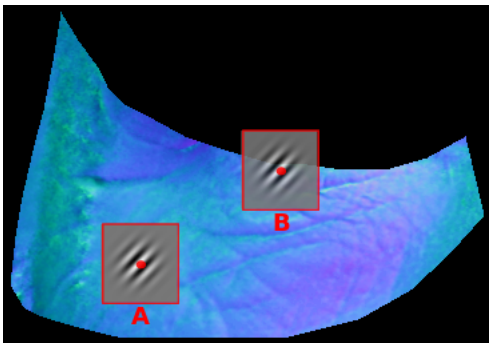


FIGURE 6.2: Issue of applying filter on irregular edges

areas of the patch. The issue can be limited by choosing small squares. The smaller the squares are, the smaller the areas left out. However, too small squares will increase the number of samples and decrease likelihood of texture disparity within a square, therefore producing an important number of not necessarily discriminative samples. So there is a trade-off to consider when choosing the sample size. We have experimented with different sized squares as presented later in this section.

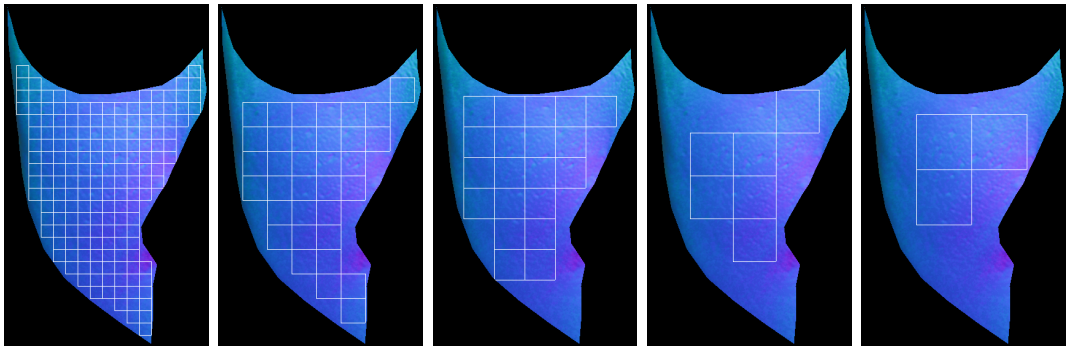


FIGURE 6.3: Example of segmentation of a frontal cheek patch with different square size

6.2.2 Baseline Implementation (BTF Image-Texton)

As described in Section 2.3.3 the BTF Image-Textons, introduced in [10, 11], are one of the most advanced and accurate representations of natural surface that accounts for 3D properties to date according to a survey in [22]. Furthermore, the photometric data, used as input for the BTF Image-Texton, and the normal map, used as input for our geometric methods, are interconvertible. Normal maps are generated form the photometric data and inversely, photometric data can be generated from normal map under a chosen lighting. In fact one of the motivation of the thesis is taking advantage of the fact that normal maps are a more compact representation of photometric data than the BTF and hence produce more compact feature vectors.

As a reminder, the method involves applying a bank of oriented filters to a BTF database. A BTF database contains images of several surfaces each of these captured under various light directions and viewpoints. This method is different to the 3D-Texton introduced in [48] in that there is no need to register the different images of a sample as the feature vectors are not constructed pixel-wise across all the images of the sample. Instead the filter responses at each pixel of each image constitutes a feature vector on its own.

As shown in Section 6.2.1 and in Chapter 5, our dataset contains multi-view data as well as photometric data. The cross-polarisation technique used during the capture allows us to obtain pure surface reflectance data (specular reflectance) under different light directions and viewpoints. Thus a part of the dataset can

be used as a BTF where each sample is captured under seven light directions and three viewpoints for the forehead or two viewpoints for the cheeks or eyelids. This yields 21 (forehead) or 14 images (cheeks or eyelids) for each sample. Figure 6.4 shows the set of 14 specular images for a cheek sample under seven light directions and two views.

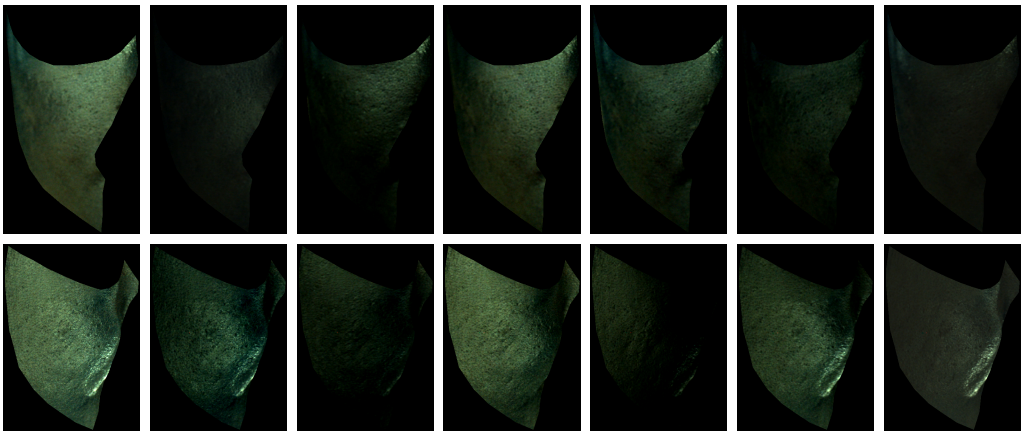


FIGURE 6.4: BTF of a cheek sample under the seven light directions and two views

Our implementation of the BTF Image-Texton consists of applying a bank of filters (with six orientations, four difference of Gaussian and four Gaussian) to all the images of each sample. This is done at three scales, which yields at each pixel a response vector of 42 elements. All the responses of all the images in the dataset are then clustered using a k-means algorithm, where K, denoting the number of cluster centres, is fixed to 200 as in [10]. Each cluster is associated with a label that corresponds to a unique texton. The histogram of textons is then computed. This represents the feature vector associated with the corresponding sample.

6.2.3 Feature Selection Method

All the texture descriptors described in Chapter 4 produce long feature vectors. The most compact of these are 128-long whereas the longest can have up to 600 components. There are many techniques that can be used to reduce these feature vectors in order to keep only the components that are the most discriminant and limit the risk of learning an over-fitted model. These techniques can be divided into two categories. The first consists of projecting the original data onto a new basis whose axis are generally chosen according to the variance distribution in the data. The second category involves ranking the relevance of each feature component. In contrast to the projection methods which produce new data, the ranking methods keep the original data.

We use in this study Support Vector Machine Ranking [32] which falls into the second category. SVM Ranking is proven to be efficient in selecting relevant feature when the dimensionality is high compared to the available number of training samples. It has been originally used successfully in gene selection for cancer classification on a dataset that contains only 62 samples but each of these containing 2000 genes which gives a ratio $feature - dimensionality/dataset - size$ over 33.

The SVM Ranking algorithm processes in a recursive way. It successively restricts the training set to the best features, then uses SVM to train on it and obtain a weight for each remaining feature. At each step, the feature with the

smallest ranking is eliminated from the training set. The process continues recursively until all the features are eliminated. In this work, We use the Weka implementation of this algorithm and chose to keep only the 64 features with the best ranking.

6.2.4 Classification Method

In this work, we use multi-layer perceptrons for training and classification. Preliminary investigations with other machine learning techniques, including Random Forests and Support Vector Machine, conducted on the 3DRFE dataset [53], indicated that the Multi-layer Perceptrons gave the best results. However as our main objective in this thesis is to propose a means of characterising 3D surface texture and study how these compare to the 2D case, instead of analysing how different Machine Learning algorithms perform on these, we only show the classification results using Multi-layer Perceptrons. A more detailed study comparing a larger range of classifiers on the proposed descriptors and the newly collected dataset is a good candidate for future work.

A Multi-layer Perceptron is a feed-forward artificial neural network. It consists of several layers of neurons. Each layer is characterised by a set of activation functions associated to each neuron it contains. Each activation function transforms a linear combination of the preceding layer's outputs and passes the results on to the next layer. These activation functions are sigmoid in the form of:

$$f(x) = \frac{1}{1 + \exp(-x)} \tag{6.1}$$

with x being a linear combination of the previous layer's output (or the input data in the case of the first layer) and have the form:

$$x = \sum_i w_i o_i \tag{6.2}$$

with w_i being the weight to learn and o_i the previous layer's output (or the input data in the case of the first layer).

Training the Multi-layer Perceptron (i.e. learning the weights w_i at each layer) consists in fixing arbitrary weights, computing the output produced by the network on the input vectors from the training set and calculating the error between these outputs and the actual labels from the training set. These errors are finally back-propagated through the network to update the weights.

In this work, we used the Weka implementation of multilayer perceptrons. We also use Weka's default as the number of hidden layers: the mean of the number of classes and the number of attributes. The output of the classifier is a discrete rating of the presence or absence of the considered skin condition and, as defined in the ground truth, is symbolic number between "1" and "5" (Section 5.4.2).

6.3 Results

In this section, we present results classifying wrinkles, large pores and acne on our collected data using all the 2D and 3D surface texture features described in this thesis. We employ two classification strategies. In the first strategy, we consider each square patch extracted using the method described in Section 6.2.1 as a full sample for training and classification. This yields a larger dataset whose size depends on the chosen patch size. We experiment with 3 different patch sizes: 20, 50 and 80. The classification results are given in Table 6.2.

In the second strategy, instead of classifying single square patches, we classify in terms of facial regions (cheek, forehead and eyelid). The sample associated with each region is computed by averaging the responses of all the square patches it contains. This produces a smaller dataset than for the first strategy and with a fixed size of $50 \times 3 = 150$ samples. However we still experiment with different square patch sizes in this strategy. Classification results are presented in Table 6.3.

For the rest of this thesis, we will refer to the first strategy as "Patch Sample" and to the second as "Region Sample".

6.3.1 Patch Sample Strategy

As stated previously, in this classification strategy, we extract square normal map patches as shown in Section 6.2.1. We experiment with three different patch sizes: 20×20, 50×50 and 80×80. Table 6.1 shows the number of extracted samples

according to the chosen patch size. For each of the 2D or 3D surface texture analysis methods presented in Chapters 3 and 4, we compute the descriptor on each square patch. Depending on the method, the resulting feature vectors are relatively long. We then run SVM Ranking, as detailed in Section 6.2.3, to rank the features according to how discriminant they are in regard to the ground truth. We retain only the first 64 features as ground truth which we then feed into a Multi-layer Perceptron with and use a 10-fold cross validation strategy for testing.

Patch Size	20×20	50×50	80×80
Nb. Samples	4302	1935	874

TABLE 6.1: Number of extracted samples per patch size

The results presented in Table 6.2 show the performances of each descriptor in terms of the F-measure [71], which represents the harmonic mean of the precision and recall:

$$F_{measure} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

We also experiment with different configurations for each of the methods. The abbreviations used in Tables 6.2 and 6.3 are as follow:

- **2D R-LBPs**: Rotation-Invariant Local Binary Patterns on albedo (Section 3.3.1)
- **2D Gabor**: Gabor filtering on albedo (Section 3.3.1)

- **3D R-LBPs**: Rotation-Invariant Local Binary Patterns on surface normals (Section 3.4)
- **3D Gabor**: Gabor filtering on surface normal (Section 3.4)
- **O. APDI**: Original Azimuthal Projection Distance (Section 4.4)
- **M-APDI**: Modified Multi-scale APDI (Section 4.4)
- **BTF Texton** Bidirectional Texture Function Texton (Section 6.2.2)
- **LOP** Local Orientation Pattern (Section 4.3)
- **Rot. Fields** Rotation Fields (Section 4.2)

TABLE 6.2: Classification results using single samples as inputs

		Wrinkles			Pores			Acne		
Features \ Sample Size		20	50	80	20	50	80	20	50	80
2D R-LBPs	$radius = 2$	.53	.59	.62	.61	.63	.62	.59	.63	.60
	$radius = 5$	.60	.67	.70	.73	.73	.72	.62	.70	.70
2D Gabor	$radius = 2$	.60	.65	.72	.58	.58	.59	.60	.62	.61
	$radius = 5$	.64	.70	.75	.71	.70	.71	.71	.73	.71
3D R-LBPs	Slant/Tilt	.75	.78	.81	.79	.81	.80	.71	.73	.72
	Tangent	.70	.73	.79	.73	.75	.73	.66	.69	.65
3D Gabor	Slant/Tilt	.78	.80	.82	.83	.85	.85	.75	.76	.77
	Tangent	.74	.78	.81	.77	.79	.79	.70	.74	.72
O. APDI	-	.62	.61	.65	.63	.60	.62	.60	.63	.63
M-APDI	$depth = 1$	.62	.65	.68	.62	.62	.64	.61	.65	.64
	$depth = 2$	.69	.70	.73	.71	.70	.72	.68	.67	.70
	$depth = 4$	.75	.78	.81	.74	.76	.73	.72	.74	.75
BTF Texton	$K = 100$	.81	.85	.88	.85	.86	.86	.86	.89	.88
	$K = 200$	**.89**	**.91**	**.93**	.87	**.90**	.86	**.90**	**.92**	**.90**
L.O.P	1st P.F.	.71	.70	.76	.63	.66	.63	.75	.79	.81
	2nd P.F.	.72	.72	.77	.79	.81	.80	.73	.78	.83
Rot. Fields	$K = 100$	.78	.80	.84	.87	.89	.87	.79	.83	.83
	$K = 200$	.82	.86	.90	**.91**	**.90**	**.92**	.84	.87	.87

The overall results show that the 3D descriptors clearly outperform the 2D descriptors. First, on comparing R-LBPs and Gabor filtering on 2D and 3D data,

both texture characterisation methods show a clear improvement when used in a 3D configuration (slant/tilt or tangent space) for the classification of both wrinkles, acne and pores. Gabor descriptors tend to give better results on classifying wrinkles whereas for pores, the Rotation-Invariant Local Binary Patterns appear to achieve better results. This is possibly due to the way that Gabor filters are likely to draw out linear structures such as wrinkles. Also the filter radius used for R-LBPs and Gabor influence the results. Overall, bigger radius values tend to yield better results.

The classification performances vary with the chosen patch size too. However this seem to depend on the skin condition being classified. The results show that for all the descriptors the performance increases with the patch size when classifying wrinkles. However this pattern does not seem to appear as regularly when classifying acne or large pores.

Further analysis of Table 6.2 shows a clear improvement of the modified Multi-scale Azimuthal Projection Distance Image over the original formulation. The M-APDI with a only a depth 1, where the sole difference from the original formulation is the introduction of a new way of computing the pixels in function of the two projection coordinates (Section 4.4.2), introduces improvement in the classification results. These improvements become even more significant as the M-APDI pyramid goes deeper.

The Local Orientation Patterns, even though not multi-scale, produce comparable results to the M-APDI. A multi-scale analysis with the LOP could be

considered by either extracting the features with variable radius values or at different resolution of the normal map. Furthermore, comparing the results yielded by the first and second proposed pattern function show clear improvement using the second pattern function over the first on classifying wrinkle and pore visibility while the first pattern function does slightly better on classifying acne. This confirms our comments on the difference in behaviour between the two proposed functions (Section 4.3.4.3): high frequency surface roughness is better captured with the second function.

The BTF Texton and our proposed Rotation Fields methods yield the highest performance rates. The BTF Texton gives somewhat better classification of wrinkle and acne than the Rotation Fields, with average improvements of 0.050 on wrinkle and 0.046 on acne. However, the Rotation Fields yield slightly better results on classifying pores visibility with an average improvement of 0.033. This can be explained by the high and low frequency separation performed in the Rotation Fields and not in the BTF Texton. Furthermore, the data needed to compute the Rotation Field (i.e. normal map) has a more compact representation. Even though it is trivial to recover surface normals from BTF data or generate BTF data from surface normals, it is more practical to store or distribute a dataset in the form of normal maps instead of BTF databases which are known to be extremely demanding in storage capacity.

6.3.2 Region Sample Strategy

As stated earlier, in this strategy, the descriptors are first computed on the square patches from each facial region (cheek, eyelid or forehead) and then averaged to produce a unique feature vector for the region. This results in fewer samples for the training than with the first strategy. This yields a total of $50 \times 3 = 150$ samples for training and classification. As in the first strategy, we employ a 10-fold cross-validation to classify the three facial regions in terms of wrinkles, acne and large pores. The results are given in terms of F-measure in Table 6.3.

This strategy is more intuitive and probably more appropriate than classifying isolated patches. The different facial regions are rated directly. One could use the first strategy and then average classification results of patches over a particular region to get a rating for that region. However this would assume that the classes (ratings) are numerical and would not be adequate for ordinal ratings.

The results show a degradation in the classification performances. The overall average decline is estimated at 0.083. However, despite this global performance decline, the relative tendencies of the classification results observed in the first strategy, according to the different descriptors or experimental setup, remain the same. The 3D texture descriptors still yield better classification results than their 2D counterparts. The BTF Texton and Rotation Fields still give the highest performances with the same pattern of the first doing better on classifying wrinkles and acne while the Rotation Fields prevail on classifying pores.

TABLE 6.3: Classification results using average samples as inputs

Features \ Sample Size		Wrinkles			Pores			Acne		
		20	50	80	20	50	80	20	50	80
2D R-LBPs	$radius = 2$	.44	.48	.54	.52	.52	.51	.50	.52	.51
	$radius = 5$	.52	.58	.60	.63	.64	.63	.53	58	.59
2D Gabor	$radius = 2$	.51	.54	.63	.50	.51	.50	.52	.53	.53
	$radius = 5$	.56	.58	.63	.61	.61	.62	.62	.61	.63
3D R-LBPs	Slant/Tilt	.66	.65	.68	.70	.69	.70	.60	.65	.63
	Tangent	.66	.65	.68	.70	.69	.70	.60	.65	.63
3D Gabor	Slant/Tilt	.66	.68	.69	.71	.76	.73	.65	.65	.64
	Tangent	.62	.64	.70	.66	.68	.67	.61	.63	.61
O. APDI	-	.48	.50	.56	.55	.51	.54	.52	.52	.51
M-APDI	$depth = 1$	.50	.53	.57	.51	.53	.53	.52	.55	.56
	$depth = 2$	.57	.58	.61	.61	.60	.62	.57	.57	.58
	$depth = 4$	.61	.70	.70	.61	.65	.62	.61	.63	.63
BTF Texton (Baseline)	$K = 100$	.70	.73	.77	.75	.74	.73	.74	.77	.76
	$K = 200$	**.80**	**.83**	**.84**	.77	**.81**	.75	**.81**	**.81**	**.80**
L.O.P	1st P.F.	.60	.61	.66	.54	.50	.55	.66	.70	.72
	2nd P.F.	.61	.61	.68	.70	.72	.70	.64	.66	.70
Rot. Fields	$K = 100$	.69	.71	.75	.77	.80	.79	.70	.72	.73
	$K = 200$	.71	.73	.82	**.80**	.79	**.81**	.74	.75	.76

The inevitable non-uniform distribution of the assessed skin condition across the different facial regions certainly accounts for the overall decrease of classification performances observed with this strategy. Figure 6.4 shows an example of square patches extracted from a cheek. It is clearly noticeable how the extracted square samples look different. This cheek is likely to be judged as wrinkly and porous by a human. In fact its average ratings according to wrinkle and pore from the psychophysical experiment are respectively 4.8 and 3.9 out of 5. Yet it is obvious from Figure 6.4 that wrinkle and visible pores are not evenly spread across the cheek and are actually visible on only few square patches. So averaging the descriptor responses of the square patches in this example will smooth out details.

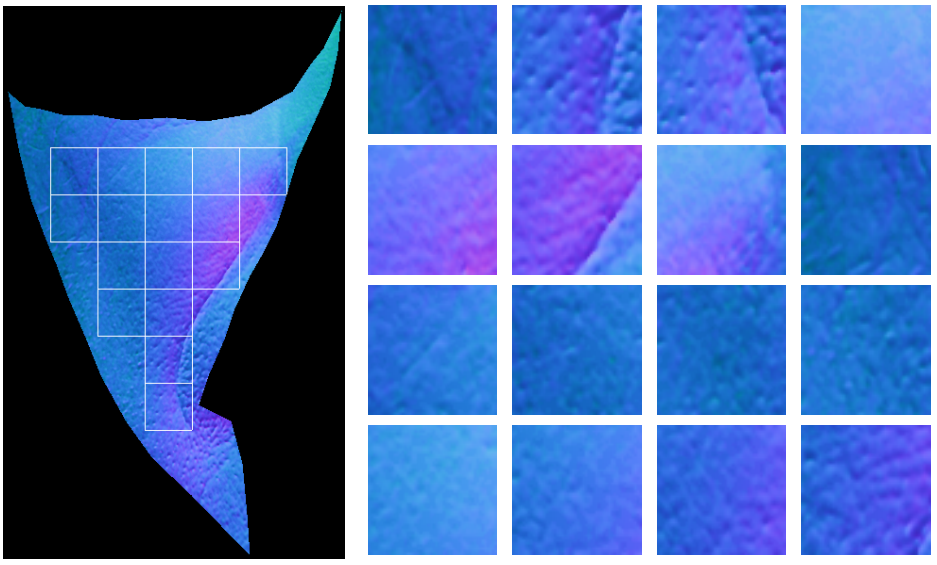


TABLE 6.4: Non-uniformity of rated skin condition over patches

This should have an effect on the first classification strategy as well. However the experimental results show that this effect is more important with the second strategy. This effect may be mitigated by a change in the setup of the ratings experiment, which could provide ground truth with more accurate localisation of the skin disruptions. Instead of presenting the participant with a whole region (cheek, eyelid or forehead) to rate, the experiment could show the entire face and let the participant freely select regions that they judge wrinkly, acneic or with visible pores. This would make analysing the inter-agreement between raters more difficult as they would not necessarily select the same regions, but we could be sure that the selected regions are those that have driven the participants rating. Another way of reducing this non-uniformity effect on the classification results would be employing an unsupervised learning method to cluster the square patches prior to averaging. But the problem would be finding an unsupervised way of deciding which clusters can be considered as outlier. Presenting the raters with square patches directly could also be considered.

6.4 Classifying Synthetic Data

For further validation of our proposed 3D surface texture descriptors, we experimented with synthetically constructed data. In Section 5.4.2, we underlined an issue inherent to the subjective nature of skin appearance rating. The rating experiment has shown that, while ratings correlate significantly between a certain number of participants, the inter-raters agreements is rather poor (Figure 5.14). This questions the validity of our ground truth and therefore the performance of the proposed descriptors. To address this, we present in this section the classification result yielded by our 3D surface texture on synthesised data where the amount of 3D texture information is mathematically introduced and therefore known a priori.

6.4.1 Synthetic Data Construction

To build the synthetic dataset, we have generated normal maps with mainly two texture patterns: spots and lines. We introduce randomness by choosing the location and size of the spots and lines at random. We control the density of the pattern distribution with a parameter d which is then used as label for classifying the generated samples.

To generate the samples, we first produce a flat depth map (all the pixel at the same depth). We then pick random pixels from the depth map by linearising the index space and using a random permutation (with the Matlab function *randperm*). The density parameter d is used as argument to the *randperm* function to

specify the number of pixels to pick. To avoid always picking the same number of pixels for all samples with the same density, we introduce another random number that modulates d. So the actual parameter to the *randperm* function is given by $max(d - \frac{d}{3}, d \times rand)$. To generate spots, we centre a Gaussian with a random standard deviation around each random pixel. For the lines, we take each random pixel as starting point and produce a line with a random length. The line is then modified with a Sine function. A Gaussian is then applied at each point of the line to emulate a depth deformation along the line. To produce the normal maps corresponding to the generated depth maps, we use the Matlab function *surfnorm*. Figures 6.5 and 6.6 show examples of generated synthetic data.

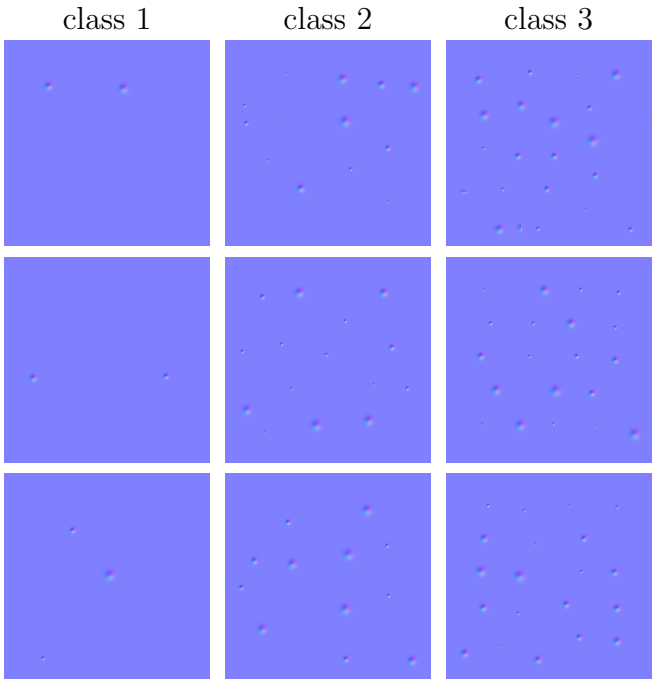


FIGURE 6.5: Example of generated synthetic data

We also produced synthetic samples with added noise. To do so, we first generate a Gaussian noise map. Then we transform the normal map from the Cartesian

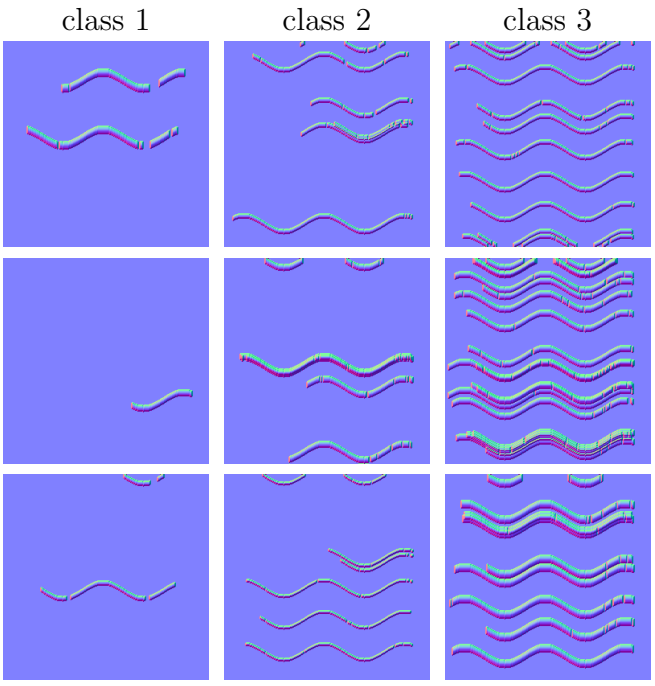


FIGURE 6.6: Example of generated synthetic data

coordinate system to the Spherical coordinate system and add the values from the noise map either to the azimuthal angle or the polar angle (the choice is random). Finally, we transform the result back to the Cartesian space. Figure 6.7 gives an illustration of the result.

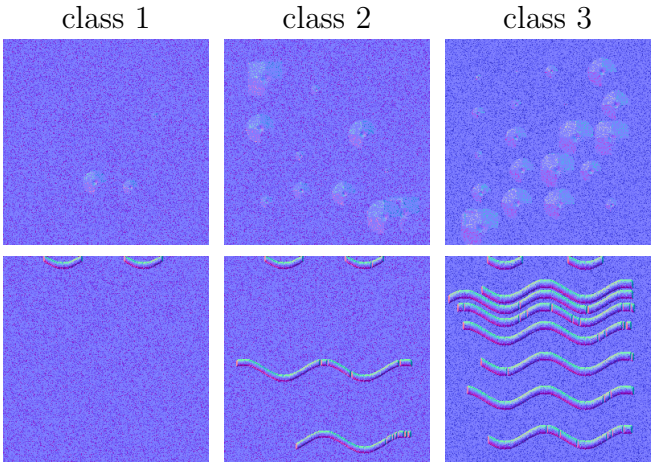


FIGURE 6.7: Example of generated synthetic data with noise

6.4.2 Classification Results

Table 6.5 shows the result of classifying the synthetic data in terms of F-measure. We have generated 3000 patches of 256×256, 1500 with lines and 1500 with spots. Each of the three classes has 500 samples. We run the three proposed 3D surface texture descriptors and classified the outputs using the same classifier and configuration as in the previous section.

Compared to the result obtained on classifying the measured data (Tables 6.2 and 6.3), the classification results in Table 6.5 generally show higher performances on the synthetic data without noise. However, augmenting the synthetic data with noise degrades significantly the classification performances. This degradation seems to be subject to the proportion of high frequency information kept by the descriptor used. The M-APDI with a depth of one and the L.O.P. with the second pattern function, which tend to keep only the high frequency information, yield the worse results on the noisy data. On the other hand the deeper level of the M-APDI and the Rotation fields tend to give better result on the noisy data as they keep both high and low frequency information through their multi-scale schemes.

TABLE 6.5: Classification results on synthetic data

Features \ Texture Type		Lines		Spots	
		no noise	noise	no noise	noise
M-APDI	$depth = 1$	.80	.50	.79	41
	$depth = 2$	.81	.66	.85	.47
	$depth = 4$	.87	.71	.88	.61
L.O.P	1st P.F.	.83	.61	.75	.63
	2nd P.F.	.85	.54	.88	.43
Rot. Fields	$K = 100$	.88	.75	.92	.69
	$K = 200$	.91	.79	.96	.71

6.5 Summary

This chapter described the experiments we conducted to compare the three 3D surface texture descriptors proposed against each other as well as against a BTF-Texton based method. The proposed 3D surface texture descriptors are extracted from the specular normal maps whereas the BTF-Texton method is applied to the gradient photometric specular albedo images. Two classification strategies are proposed. In the first one, the classification samples are individual square patches extracted from the facial regions whereas in the second strategy, the samples are constructed by averaging the responses of all patches extracted from the same facial region (cheek, forehead or eyelid). Support Vector Machine Ranking is used for feature selection and a Multilayer Perceptron is used for classification.

The overall results show clear improvement when 3D texture descriptors are used. Despite outperforming traditional 2D descriptors, the Local Orientation Patterns and Multi-scale APDI are surpassed by the BTF-Texton based method and the Rotation Fields which, together, yield the highest classification performances. While the BTF-Texton does slightly better on classifying wrinkles and acne, the Rotation Fields achieve better results on classifying Pores. This can be explained by the high/low frequency separation present in the Rotation Field method. Another advantage of the Rotation Map over the BTF-Texton resides in the fact that the data needed in the former (surface normals) have more compact representation than data required to build the BTF. Whilst normal maps could be easily recovered from BTF database and vice-versa, it is clear that, in terms of

storage, processing and/or distribution, it is more practical to work with normal maps than BTF databases.

Chapter 7

Discussions and Conclusions

This thesis has explored new methods of characterising the 3D nature of surface texture and has applied these to facial skin texture analysis and synthesis. For a long time, texture has been dealt with as a 2D problem. However, in the late nineties, new studies emerged to tackle the issue, taking into account changes in apparent texture subject to imaging conditions, which is inherent to the 3D nature of rough surface texture. These pioneering studies include [12] and [48]. Most of these studies work on BTF data which are constructed by capturing the surface texture under various viewpoints and light directions. These BTF dataset are known to be very complex to capture and require considerable storage space (up to hundreds of gigabytes). Also these methods are image-based as the intrinsic geometry of the surface is unknown.

Today, new methods of efficiently capturing 3D surface texture have emerged. Photometric stereo allows capturing not only surface micro-geometry, but also

the surface reflectance properties. Thus new methods of extracting texture descriptors from these newly available data are desirable. In this thesis, we use photometric spherical gradient methods, introduced in [53], to capture 3D face models along with highly detailed surface normals and reflectance properties (Chapter 5). The captured dataset consists of 50 3D face models of subjects presenting different skin conditions, along with specular and diffuse normal maps and albedos.

In contrast to image-based methods which use BTF data, the surface texture descriptors proposed in this thesis operate directly on the captured surface micro-geometry in the form of dense surface normals. Three novel methods of characterising 3D surface texture are proposed. The performances of these are evaluated on classifying common skin conditions (wrinkles, large pores, acne) and compared against state-of-the-art methods represented by a BTF-Texton based approach. We have also compared the performances of traditional two-dimensional texture measures (LBPs and Gabor filters bank) and very simplistic extensions of these to the 3D space.

7.1 2D Vs 3D Surface Texture

We presented in Chapter 3 a comparative study on classifying skin conditions using traditional 2D texture descriptors (LBPs and Gabor filter banks) and their 3D counterparts. The 2D descriptors are applied on the measured specular albedo while the 3D counterparts are constructed by extending these 2D descriptors to

the slant/tilt and tangent spaces of measured surface normals. The classification results presented in Tables 6.2 and 6.3 show improvement in the 3D configuration. While the extensions of the 2D descriptors to the 3D space proposed are rather simplistic, these results give a first appreciation of the value added by considering the 3D aspect of surface texture.

These improvements are inherent to the geometrical micro-structure of skin texture as with a number of other natural materials. As stated earlier, the apparent texture of natural surfaces is strongly subject to the imaging conditions. A surface captured under different light conditions and viewpoints will yield different appearances in texture. An image captured in a fixed light/view environment will convey only the component of the texture responsive to that particular setup, potentially leaving out considerable amounts of textural information. This is as important as the captured surface exhibit roughness. Thus it is only natural that characterising the texture of natural material such as human skin yields better performances in a three-dimensional configuration.

7.2 Proposed 3D Surface Texture Descriptors

Three methods of characterising 3D surface texture from dense normals fields were proposed in this thesis and are briefly discussed bellow.

Rotation Field Pyramids

This method separates the surface normals' high frequency components from the low frequency components. The separation can be achieved at different scales yielding a multi-resolution pyramid. At each level of the pyramid, the normals field is smoothed (using Euclidean or Geodesic distance) to produce the low frequency components. Then the high frequencies are extracted by calculating the rotation fields that separate the smoothed normals from the original ones (at that scale). Features are extracted by calculating in a neighbourhood the distances between all the rotations in that neighbourhood to the central one. K-means are then applied to quantise the results before extracting local histograms that constitute the feature vectors. Of the three methods proposed in this thesis, this one achieved the best classification performance. We also showed that Rotation Field Pyramids can be used to synthesise/transfer highly detailed 3D surface texture to a new face which can be used to synthesise ageing or other skin conditions such as acne or large pores.

Local Orientation Patterns

This method is an extension of the notion of Texture Spectrum introduced in [85] and which has influenced the Local Binary Patterns proposed in [62]. The main idea is to compare the normals orientation in a neighbourhood with the central normal. Two methods of comparing normal orientation are proposed. The first one compares the normals' azimuthal and polar angles and the second applies a

threshold to their dot product. Our experimental results have shown that the first method is more appropriate for extracting high frequency texture whereas the second give better performances while extracting coarser structures.

Multi-scale Azimuthal Projection Distance

This extends the Azimuthal Projection Distance Image (APDI) proposed in [72]. Three additions to the original method are proposed which significantly improve the classification results from the original formulation. First, instead of using a global fixed projection axis, a local mean is computed using an Euclidean or geodesic metric. Second, the APDI is modified to introduce a term that is related to normal azimuthal orientation. Finally a Multi-resolution analysis scheme is conducted so that skin disruptions at different scales can be captured.

7.3 Performance Evaluation

The proposed 3D surface texture descriptors are evaluated on the collected dataset by classifying wrinkles, large pores and acne from dense normal fields measured using gradient photometric stereo. The three methods are compared against each other and against an implemented baseline based on BTF-Texton [10, 11] which is to date the most advanced representation of 3D surface textural information according to [22].

The experiments show that, of the three proposed methods, the Rotation Fields produce the best classification results with average F-measures of 0.86, 0.91 and

0.86 on classifying respectively wrinkles, pores and acne. The BTF-Texton based method performs better than the Rotation Fields on classifying wrinkles and and acne with respectively F-measures of 0.91 and 0.90. However on classifying pores the Rotation Fields give somewhat better result with 0.91 against 0.87. This suggests that the Rotation Fields are more efficient on characterizing high frequency conditions such as pores whereas lower frequency conditions such as wrinkle and acne are somewhat better classified using the BTF-Texton. As stated in Chapter 6, this can be explained by the high and low frequency separation performed in the Rotation Fields method and not in the BTF Texton method. If, for now, the BTF-Texton based method performs better than the Rotation field on classifying low frequency condition with average improvements of 0.050 and 0.046 (in terms of F-measure) respectively for wrinkle and pores, the other benefits of the Rotation fields should be considered as well. These include a more compact representation of both feature vectors and datasets and the ability to take advantage of recent advances made in 3D surface capture techniques.

7.4 Future Work and Outlook

7.4.1 3D Surface Texture Data

The 3D surface texture descriptors proposed operate on dense surface normal fields. The normal fields used in this thesis are captured using a cross-polarised photometric stereo method. While the experiments conducted in this study show clear improvements when surface texture is considered as a 3D problem, most

existing applications and available datasets are 2D oriented. Thus, these improvements will not benefit a large range of applications. [48] have tried to tackle this issue when they proposed the 3D texton. They have proposed a method of using their 3D texton vocabulary to recognise texture from single 2D images with unknown view/illumination conditions. Their approach consists of initialising input labels by randomly assigning each pixel to a 3D texton and iteratively computing the probability of that assignment and eventually updating it until the process converges. Similar approaches could be used to predict the 3D structure of a surface from a single image. Moreover, the problem of predicting surface 3D properties from single image with unknown view/illumination conditions generate more and more interest in the computer vision community. A growing number of studies try to tackle this issue by considering it as a statistical inference problem that can be solved using optimisation techniques [1, 50, 81].

7.4.2 Quantifying Skin Condition

In this study, we have tried to build a system that is able to automatically assess certain skin conditions. This is already not a trivial task for humans. The way people perceive skin appearance is rather subjective. It is difficult, if not impossible, to obtain from different people a unanimous judgement on skin conditions. The rating experiment we have conducted in this thesis is a proof of that. We presented in Section 5.4.2 the outcomes of our experimental skin condition human rating and Figure 5.14 show poor agreement between participants even if there

is significant correlation between some of them. Therefore, we have used correlation instead of agreement as a criteria of eliminating rating outliers even though the latter would be more appropriate. So we think that future work should include a more psychophysics-oriented experiment aiming at studying how human perceive skin condition and maybe propose a quantitative metric that could be used in automated assessment systems.

7.4.3 Classification Methods

The main objective of this thesis were to propose new methods of characterising 3D surface texture for skin condition classification. While our focus were more directed towards the texture characterisation part, we believe that further experiments comparing different classification methods on the proposed 3D texture descriptors would be of significant value. As stated in Section 6.2.4, our choice of using the Multilayer Perceptron were driven by a primitive experiment on a limited dataset where we have considered the Multilayer Perceptron, the Support Vector Machines and the Random Forest. But this needs to be confirmed by a more appropriate experiment on the collected dataset.

7.4.4 Towards Automated Skin Health Assessment

The techniques developed in this thesis, although experimental, could be a step towards an entirely automated system for analysing and assessing skin health. Such a tool would be useful to number of applications. Dermatologists could use

it to assist skin conditions diagnosis. It could also be used in the cosmetic industry for evaluating products' effects on skin. Even consumer level applications could be considered. However this would require a certain number of improvements and adaptations. First, the capture system would need to be simplified and made more accessible. The lightstage used in this study has a diameter of 1.58m, 41 LEDs and 7 DSLRs.

Another improvement that would help towards a fully automated system would be a face landmarking model that would automatically segment the 3D face model into different regions (cheek, forehead, eye-corner, chin etc.). As stated in Section 5.3.2.4, this segmentation is done manually for now by placing 2D landmarks in the texture space before back-projecting them to the mesh. A more automated alternative would be to use a 2D facial landmarks detector such as [43, 70, 83] on each of the view before projecting them onto the geometry. There are also 3D face landmarking algorithms that could be tried [15, 78].

Bibliography

[1] Barron, J. and Malik, J. (2015). Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37:1670–1687.

[2] Bonet, J. D. (1997). Multiresolution sampling procedure for analysis and synthesis of texture images. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, pages 361–368.

[3] Breitenstein, M., Kuettel, D., Weise, T., Gool, L. V., and Pfister, H. (2008). Real-time face pose estimation from single range images. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8.

[4] Brox, T., Bruhn, A., Papenberg, N., and Weickert, J. (2004). High accuracy optical flow estimation based on a theory for warping. In *European Conference on Computer Vision (ECCV)*, volume 3024, pages 25–36.

[5] Brox, T. and Malik, J. (2011). Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:500–513.

[6] Bruhn, A., Weickert, J., and Schnörr, C. (2005). Lucas/kanade meets horn/schunck: combining local and global optic flow methods. *International Journal of Computer Vision*, 61:211–231.

[7] Caputo, B., Hayman, E., and Mallikarjuna, P. (2005). Class-specific material categorisation. In *10th IEEE International Conference on Computer Vision, 2005. ICCV 2005*, volume 2, pages 1597–1604.

[8] Cohen, F., Fan, Z., and Patel, M. (1991). Classification of rotated and scaled textured images using gaussian Markov Random Field models. *IEEE Transaction Pattern Analysis Machine Intelligence*, 13:192–202.

[9] Cook, R. and Torrance, K. (1982). A reflectance model for computer graphics. *ACM Transactions Graphics*, 1:7–24.

[10] Cula, O. and Dana, K. (2004). 3D Texture recognition using bidirectional feature histograms. *International Journal of Computer Vision*, 59:2004.

[11] Cula, O., Dana, K., Murphy, F., and Rao, B. (2004). Bidirectional imaging and modeling of skin texture. *IEEE Transactions on Biomedical Engineering*, 51:2148–2159.

[12] Dana, K., Ginneken, B. V., Nayar, S., and Koenderink, J. (1999). Reflectance and texture of real-world surfaces. *ACM Transaction Graphics*, 18:1–34.

[13] Daugman, J. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2:1160–1169.

[14] Dessein, A., Smith, W., Wilson, R., and Hancock, E. (2014). Seamless texture stitching on a 3d mesh by poisson blending in patches. In *IEEE International Conference on Image Processing (ICIP)*, pages 2031–2035.

[15] Dibeklioglu, H., Salah, A., and Akarun, L. (2008). 3d facial landmarking under expression, pose, and occlusion variations. In *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6.

[16] Ding, Y., Smith, L., Smith, M., Sun, J., and Warr, R. (2009). Obtaining malignant melanoma indicators through statistical analysis of 3D skin surface disruptions. *Skin Research and Technology*, 15:262–270.

[17] Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in sobolev spaces. In *Constructive theory of functions of several variables*, volume 571 of *Lecture Notes in Mathematics*, pages 85–100.

[18] Duda, R. and Hart, P. (1972). Use of the Hough transformation to detect lines and curves in pictures. *Communication of ACM*, 15:11–15.

[19] Efros, A. and Freeman, W. (2001). Image quilting for texture synthesis and transfer. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pages 341–346.

[20] Efros, A. and Leung, T. (1999). Texture synthesis by non-parametric sampling. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999.*, volume 2, pages 1033–1038.

[21] Eichmann, G. and Kasparis, T. (1988). Topologically invariant texture descriptors. *Computer Vision Graphics Image Processing*, 41:267–281.

[22] Filip, J. and Haindl, M. (2009). Bidirectional texture function modeling: a state of the art survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:1921–1940.

[23] Fish, P., Round, A., and Duller, A. (2000). Lesion classification using skin patterning. *Skin Research and Technology*, 6:183–192.

[24] Fish, P. and She, Z. (2002). Skin lesion differentiation using skin line direction. In *Proceedings of Medical Image Understanding and Analysis*.

[25] Fish, P. and She, Z. (2003). Analysis of skin line pattern for lesion classification. *Skin Research and Technology*, 9:73–80.

[26] Fogel, I. and Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics*, 61:103–113.

[27] Freeman, W. and Adelson, E. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:891–906.

[28] Funkhouser, T., Golovinskiy, A., Matusik, W., Pfister, H., and Rusinkiewicz, S. (2006). A statistical model for synthesis of detailed facial geometry. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*.

[29] Goyal, R., Goh, W., Mital, D., and Chan, K. (1995). Scale and rotation invariant texture analysis based on structural property. In *Proceedings of the 1995 IEEE IECON 21st International Conference on Industrial Electronics, Control, and Instrumentation*, volume 2, pages 1290–1294.

[30] Graham, P., Tunwattanapong, B., Busch, J., Yu, X., Jones, A., Debevec, P., and Ghosh, A. (2012). Measurement-based synthesis of facial microgeometry. In *ACM SIGGRAPH*.

[31] Greenspan, H., Belongie, S., Goodman, R., and Perona, P. (1994). Rotation invariant texture recognition using a steerable pyramid. In *Proceedings of the 12th International Conference on Pattern Recognition*, volume 2, pages 162–167.

[32] Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46:389–422.

[33] Han, J. and Perlin, K. (2003). Measuring bidirectional texture reflectance with a kaleidoscope. *ACM Transaction Graphics*, 22:741–748.

[34] Hanrahan, P., Jensen, H., Marschner, S., and Levoy, M. (2001). A practical model for subsurface light transport. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*.

[35] Haralick, R. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804.

[36] Heeger, D. and Bergen, J. (1995). Pyramid-based texture analysis/synthesis. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, pages 229–238.

[37] Hertzmann, A., Jacobs, C., Oliver, N., Curless, B., and Salesin, D. (2001). Image analogies. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pages 327–340.

[38] Hormann, K. and Agathos, A. (2001). The point in polygon problem for arbitrary polygons. *Computer Geometry Theory and Applications*, 20:131–144.

[39] Jain, A. and Farrokhnia, F. (1990). Unsupervised texture segmentation using Gabor filter . In *Systems, Man and Cybernetics, 1990. Conference Proceedings., IEEE International Conference on*, pages 14–19.

[40] Jia-Lin, C. and Kundu, A. (1995). Unsupervised texture segmentation using multichannel decomposition and hidden Markov models. *IEEE Transactions on Image Processing*, 4:603–619.

[41] Jing, W. and Dana, K. (2006). Relief texture from specularities. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28:446–457.

[42] Jirí, T. and Jirí, M. (2010). Extended set of local binary patterns for rapid object detection. In *Proceedings of the Computer Vision Winter Workshop*.

[43] Kazemi, V. and Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1867–1874.

[44] Koh, J., Choi, K., Kim, S., Baek, J., Kang, S., and Boo, Y. (2012). Cosmetic efficacy evaluation of an anti-acne cream using the 3D image analysis system. *Skin Research and Technology*, 18:192–198.

[45] Kopf, R., Friedman, D., and Rigel, A. (1985). Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. *A Cancer Journal for Clinicians*, 35:130–151.

[46] Koshimizu, H., Hayashi, J., and Yasuoto, H. (2001). Method for estimating and modeling age and gender using facial image processing. In *7th international conference Virtual Systems and Multimedia*.

[47] Kwatra, V., Schödl, A., Essa, I., Turk, G., and Bobick, A. (2003). Graphcut textures: image and video synthesis using graph cuts. In *ACM SIGGRAPH*, pages 277–286.

[48] Leung, T. and Malik, J. (2001). Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal Computer Vision*, 43:29–44.

[49] Liu, C., Yang, G., and Gu, J. (2013). Learning discriminative illumination and filters for raw material classification with optimal projections of bidirectional texture functions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1430–1437.

[50] Liu, X., Jiang, L., Wong, T., and Fu, C. (2012). Statistical invariance for texture synthesis. *IEEE Transactions on Visualization and Computer Graphics*, 18:1836–1848.

[51] Lloyd, S. (2006). Least squares quantization in pcm. *IEEE Transaction Information Theory*, 28(2):129–137.

[52] Ma, W. (2008). *A framework for capture and synthesis of high resolution facial geometry and performance*. PhD thesis, National Taiwan University.

[53] Ma, W., Hawkins, T., Peers, P., Chabert, C., Weiss, M., and Debevec, P. (2007). Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Eurographics Symposium on Rendering*.

[54] McDaniel, T. and Panchanathan, S. (2007). Perceptual surface roughness classification of 3D textures using Support Vector Machines. In *IEEE International Workshop on Haptic, Audio and Visual Environments and Games, 2007. HAVE 2007.*, pages 154–159.

[55] Miyamoto, Y., Shirazi, M., and Uehara, K. (2000). Texture analysis and classification using bottom-up tree-structured wavelet transform. In *PRICAI 2000 Topics in Artificial Intelligence*, volume 1886, pages 802–802.

[56] Müller, G., Bendels, G., and Klein, R. (2005a). Rapid synchronous acquisition of geometry and appearance of cultural heritage artefacts. In *Proceedings of the 6th International Conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage*, pages 13–20. Eurographics Association.

[57] Müller, G., Meseth, J., Sattler, M., Sarlette, R., and Klein, R. (2005b). Acquisition, synthesis, and rendering of bidirectional texture functions. *Computer Graphics Forum*, 24:83–109.

[58] Nabiyev, V. and Gunay, A. (2008). Automatic age classification with LBP. In *23rd International Symposium Computer and Information Sciences*.

[59] Nagano, K., Fyffe, G., Alexander, O., Barbic, J., Li, H., Ghosh, A., and Debevec, P. (2015). Skin microstructure deformation with displacement map convolution. *ACM Transaction Graphics*, 34:1–10.

[60] Nehab, D., Rusinkiewicz, S., Davis, J., and Ramamoorthi, R. (2005). Efficiently combining positions and normals for precise 3D geometry. *ACM Transactions on Graphics*, 24(3).

[61] Neubeck, A., Zalesny, A., and Gool, L. V. (2004). Viewpoint consistent texture synthesis. In *2nd International Symposium on 3D Data Processing, Visualization and Transmission*, pages 388–395.

[62] Ojala, T., Pietikäinen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29‘(1):51–59.

[63] Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 24:971–987.

[64] Pennec, X. (1996). *L'incertitude dans les problèmes de reconnaissance et de recalage – Applications en imagerie médicale et biologie moléculaire*. PhD thesis, Ecole Polytechnique X.

[65] Pennec, X. (2009). Statistical computing on manifolds: From riemannian geometry to computational anatomy. In Nielsen, F., editor, *Emerging Trends in Visual Computing*, volume 5416 of *Lecture Notes in Computer Science*, pages 347–386.

[66] Perronnin, F., Dance, C., Csurka, G., and Bressan, M. (2006). Adapted vocabularies for generic visual categorization. In *European Conference of Computer Vision (ECCV)*, volume 3954, pages 464–475.

[67] Pietikäinen, M., Ojala, T., and Xu, Z. (2000). Rotation-invariant texture classification using feature distributions. *Pattern Recognition*, 33:43–52.

[68] Porter, R. and Canagarajah, C. (1997). Robust rotation-invariant texture classification: wavelet, Gabor filter and GMRF based schemes. *IEEE Proceedings - Vision, Image and Signal Processing*, 144:180–188.

[69] Portilla, J. and Simoncelli, E. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal Computer Vision*, 40:49–70.

[70] Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2879–2886.

[71] Rijsbergen, C. (1979). *Information retrieval*. Butterworth-Heinemann, Newton, MA, USA, 2nd edition.

[72] Sandbach, G., Zafeiriou, S., and Pantic, M. (2012). Binary pattern analysis for 3D facial action unit detection. In *Proceedings of the British Machine Vision Conference*.

[73] Savran, A., Alyuz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., B., B. S., and Akarun, L. (2008). Biometrics and identity management. chapter Bosphorus Database for 3D Face Analysis, pages 47–56.

[74] Schmid, C. (2001). Constructing models for content-based image retrieval. In *IEEE Conference of Computer Vision and Pattern Recognition*, volume 2, pages 39–45.

[75] Seck, A., Dee, H., and Tiddeman, B. (2014). 3D facial skin texture analysis using geometric descriptors. In *22nd International Conference on Pattern Recognition (ICPR)*, pages 1126–1131.

[76] Shu, L. and Chung, A. (2007). Texture classification by using advanced local binary patterns and spatial distribution of dominant patterns. In *IEEE International Conference on Acoustics, Speech and Signal Processing.*, volume 1, pages 1221–1224.

[77] Smith, M., Anwar, S., and Smith, L. (2011). 3D texture analysis using co-occurrence matrix feature for classification. In *Fourth York Doctoral Symposium on Computer Science*.

[78] Song, M., Tao, D., Sun, S., Chen, C., and Maybank, S. (2014). Robust 3d face landmark localization based on local coordinate coding. *Image Processing, IEEE Transactions on*, 23:5108–5122.

[79] Stratou, G., Ghosh, A., Debevec, P., and Morency, L. (2011). Effect of illumination on automatic expression recognition: A novel 3D relightable facial database . *IEEE International Conference onAutomatic Face & Gesture Recognition and Workshops*, pages 611–618.

[80] Suen, P. and Healey, G. (2000). The analysis and recognition of real-world textures in three dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:491–503.

[81] Tappen, M., Freeman, W., and Adelson, E. (2005). Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1459–1472.

[82] Tiddeman, B., Duffy, N., and Rabey, G. (2001). A general method for overlap control in image warping. *Computers and Graphics*, 25:59–66.

[83] Valstar, M., Martinez, B., Binefa, X., and Pantic, M. (2010). Facial point detection using boosted regression and graph models. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2729–2736.

[84] Varma, M. and Zisserman, A. (2002). Classifying images of materials: Achieving viewpoint and illumination independence. In *Proceedings of the 7th European Conference on Computer Vision-Part III*, pages 255–271.

[85] Wang, L. and He, D. (1990). Texture classification using texture spectrum. *Pattern Recognition*, 23:905–910.

[86] Warr, R., Zhou, Y., Smith, M., and Smith, L. (2010). Using 3D differential forms to characterize a pigmented lesion in vivo. *Skin Research and Technology*, 16:77–84.

[87] Wei, L. and Levoy, M. (2000). Fast texture synthesis using tree-structured vector quantization. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pages 479–488.

[88] Weinmann, M., Gall, J., and Klein, R. (2014). Material classification based on training data synthesized using a btf database. In *European Conference on Computer Vision (ECCV)*, volume 8691, pages 156–171.

[89] Wen-Rong, W. and Shieh-Chung, W. (1996). Rotation and gray-scale transform-invariant texture classification using spiral resampling, subband decomposition, and hidden Markov model. *Image Processing, IEEE Transactions on*, 5:1423–1434.

[90] Weyrich, T., Matusik, W., Ngan, A., and Gross, M. (2006). Analysis of human faces using a measurement-based skin reflectance model. *ACM Transaction on Graphics*, 25:1013–1024.

[91] Wilson, C., Ghosh, A., Peers, P., Chiang, J., Busch, J., and Debevec, P. (2010). Temporal upsampling of performance geometry using photometric alignment. *ACM Transaction on Graphics*, 29:1–11.

[92] Winn, J., Criminisi, A., and Minka, T. (2005). Object categorization by learned universal visual dictionary. In *Tenth IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1800–1807.

[93] Woodham, R. (1989). Shape from shading. chapter Photometric Method for Determining Surface Orientation from Multiple Images, pages 513–531. MIT Press, Cambridge, MA, USA.

[94] Yin, L., Wei, X., Sun, Y., Wang, J., and Rosato, M. (2006). A 3d facial expression database for facial behavior research. In *Proceedings of the 7th*

International Conference on Automatic Face and Gesture Recognition, pages 211–216.

[95] Zafeiriou, S., Hansen, M., Atkinson, G., Argyriou, V., Petrou, M., Smith, M., and Smith, L. (2011). The photoface database. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 132–139.

[96] Zhang, J. and Tan, T. (2002). Brief review of invariant texture analysis methods. *Pattern Recognition*, 35:735–747.

[97] Zhang, L., Chu, R., Xiang, S., Liao, S., and Li, S. (2007). Face detection based on multi-block LBP representation. In *Proceedings of the 2007 International Conference on Advances in Biometrics*, pages 11–18.